Eucalyptus genomics: Linkage mapping, QTL analysis and population genomic studies

Corey James Hudson

B. Sc. (Hons)

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Plant Science, University of Tasmania, January 2012

Declaration

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of the my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Corey James Hudson

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Corey James Hudson

The research associated with this thesis abides by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Corey James Hudson

Abstract

This thesis reports on genomic studies conducted in five of the most commercially important species of *Eucalyptus*, all from subgenus *Symphyomyrtus*; *E. globulus*, *E. nitens* (section *Maidenaria*), *E. grandis*, *E. urophylla* (section *Latoangulatae*) and *E. camaldulensis* (section *Exsertaria*). These studies utilised Diversity Arrays Technology (DArT), a microarray-based genotyping system which provides high-throughput genome-wide genotyping. DArT markers, and in particular linkage maps constructed with these markers, formed the basis for most of the research conducted in this thesis. The DArT markers are highly transferable across species which facilitates the transfer of information between studies and as sequences are available for most of the 7680 markers contained on the eucalypt genotyping array it is possible to place these markers on the recently released *E. grandis* genome sequence.

A 1060 DArT and microsatellite marker linkage map was constructed in a large *E. globulus* dwarf x tall ecotype out-crossed F_2 family (n = 503) in Chapter 1. This is the highest density linkage map produced for *E. globulus* and together with other high density DArT maps constructed in *E. grandis* x *E. urophylla* it was used to assess genome similarity (e.g. synteny and colinearity) between species through comparative mapping. Despite the detection of two small putative translocations or duplications in the inter-sectional comparison, these species showed very high synteny and colinearity overall. This finding showed that it is possible to link the economically important *E. globulus* to the *E. grandis* genome sequence.

The feasibility of constructing a multi-species composite linkage map for *Symphyomyrtus* was explored in Chapter 2. Seven independently constructed linkage maps (built in either *E. globulus* or *E. grandis* x *E. urophylla* mapping families) were integrated into a single map containing a total of 4135 markers (3909 DArT markers, 218 microsatellite markers and 8 candidate genes). This composite map will, (1) serve as a valuable reference map for *Symphyomyrtus* and related species, and (2) aid comparative research by enabling the relative positions of markers and quantitative trait loci (QTL) from different studies to be determined more easily.

Quantitative trait loci analyses were used to examine the genetic architecture underlying differences between dwarf and tall *E. globulus* ecotypes in Chapter 3; with a particular focus on vegetative and reproductive phase-change traits. A total of 22 significant and

11 suggestive QTL were detected for the nine traits examined. A highly significant QTL was detected for vegetative phase-change (LOD 103, 62.8% phenotypic variation explained). A micro-RNA gene known to regulate vegetative phase-change was found near the peak of this QTL. It is hypothesised that a mutation in, or a deletion of, this gene may be responsible for the heterochronic evolution of early phase-change in dwarf ecotype populations and that this has allowed *E. globulus* to inhabit the exposed coastal environments in which the dwarf ecotype occurs.

In Chapter 4, population genomic analyses were conducted using range-wide samples (84 to 93 individuals per species) of *E. globulus*, *E. nitens*, *E. grandis*, *E. urophylla* and *E. camaldulensis*. Genetic diversity parameters were calculated using a common set of 2207 markers. In each of ten pair-wise species comparisons, markers under putative directional selection were identified using F_{ST} outlier analyses. These outlier markers along with genetic diversity estimates were mapped on the composite map (Chapter 2) to produce a 'genomic atlas'. This identified several genomic 'hot-spots' of species molecular differentiation and together with other genetic resources, such as the *E. grandis* genome sequence and QTL positional information, offers the potential to identify candidate genes underlying traits associated with species adaptation and speciation.

Acknowledgements

This thesis would not have been possible without the valuable assistance provided by many people. Foremost, I would like to thank my supervisors, René Vaillancourt, Brad Potts and Jules Freeman, for their excellent guidance, support and encouragement. Rebecca Jones, Dorothy Steane, Valérie Hecht and Jim Weller are acknowledged for sharing their expertise. Valuable field and or laboratory assistance was provided by James Marthick, Sascha Wise, Paul Tilyard, Justin Bloomfield and Adam Smolenski (CSL).

I thank Dario Grattapaglia for providing materials, access to equipment, and the opportunity to conduct research in his laboratory at Embrapa Genetic Resources and Biotechnology, Brazil. A special thanks to Danielle Faria for her laboratory assistance and support during this time. Marilia Pappas and Juliana Sena are acknowledged for their laboratory assistance. Carolina Sansaloni and César Petroli are thanked for providing linkage mapping and DArT marker information. Samuel Paiva is also thanked for his support during my Brazil trip.

Zander Myburg and Anand Kullan from the University of Pretoria are thanked for their contributions to the comparative mapping and composite map projects. They are also acknowledged along with Geoff Galloway at Sappi (South Africa) and Kitt Payn at Mondi (South Africa) for their assistance and contribution of *E. grandis* and *E. urophylla* DNA, respectively.

Andrzej Kilian and Frank Dettering at DArT Pty Ltd, Canberra are thanked for assisting with DArT genotyping, linkage mapping and QTL analyses. Kelsey Joyce at Gunns Ltd., Dean Williams at Forestry Tasmania and Sandra Hetherington at Norske Skog provided access to field trials and materials. Simon Southerton, Penny Butcher and Charlie Bell at CSIRO Plant Industries, Canberra are thanked for their assistance and contribution of *E. camaldulensis* DNA.

I would like to acknowledge the late Sir Harold Cuthbertson and Lady Cuthbertson whose generosity provided the stipend that allowed me to undertake this work, the Australian Research Council for providing funding to this project (DP0770506 & DP110101621) and the Co-operative Research Centre for Forestry for in-kind support.

Finally, special thanks to my parents, Ron and Wendy Hudson, for their support throughout my many years of study.

Publications arising from this project

Journal Publications

Hudson CJ, Kullan ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM, Vaillancourt RE (2011) High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. *Tree Genetics & Genomes* (http://dx.doi.org/10.1007/s11295-011-0444-9)

Conference proceedings

Freeman JS, Potts BM, Downes GM, Thavamanikumar S, Pilbeam DJ, **Hudson CJ**, Vaillancourt RE (2011) QTL analysis for growth and wood properties across multiple pedigrees and sites in Eucalyptus globulus**.** *BMC Proceedings* **5** (Suppl 7).

Hudson CJ, Kumar KAR, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Potts BM, Myburg AA and Vaillancourt RE (2010). Comparative mapping with high marker density reveals high synteny and colinearity among *Eucalyptus* genomes. Oral presentation at Genetics Society of AustralAsia (GSA) Conference, Canberra, Australia 4-8th July 2010. Abstract page 14.

Hudson CJ, Kumar KAR, Freeman JS, Potts BM, Myburg AA and Vaillancourt RE (2010) High genome homology between *Eucalyptus* species. Poster presentation at CRC for Forestry Annual Conference, Fremantle, Western Australia, 17-20th May 2010.

Hudson CJ, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Potts BM and Vaillancourt RE (2010). High-density linkage map in *Eucalyptus globulus* constructed with 500+ F_2 progenies using Diversity Array Technology (DArT) and microsatellite markers. Poster presentation at Plant and Animal Genome (PAG) 18 Conference, 9-13th January 2010, San Diego, California. Abstract page 212.

Steane, DA, **Hudson, CJ**, Nicolle, D, Sansaloni, CP, Petroli, CD, Carling, J, Kilian, A, Myburg, AA, Grattapaglia, D, Vaillancourt, RE (2011) DArT markers provide a powerful tool for population genetic analysis and phylogeny reconstruction in

Eucalyptus (*Myrtaceae*). Poster presentation at the 28th International Botanical Congress, 23rd - 30th July 2011, Melbourne, Australia. Poster abstract P0423

Vaillancourt, RE, **Hudson, CJ**, Kumar, AR, Myburg, AA, Faria, DA, Grattapaglia, D, Kilian, A (2011) Comparative mapping in *Eucalyptus* genomes. Oral presentation by RE Vaillancourt at the 28th International Botanical Congress, 23rd - 30th July 2011, Melbourne, Australia. Abstract page 98.

Vaillancourt RE, McKinnon GE, Freeman J, Jones RC, Steane DA, **Hudson CJ**, Potts BM (2010). Evolutionary genetic studies in the Tasmanian species of *Eucalyptus*. Invited lecture presented by RE Vaillancourt for the symposium 'Plant genetic insights in the southern hemisphere: from Gondwana till present' at the VI Southern Connection Congress, Bariloche, Argentina, 15-19th February 2010.

Vaillancourt RE, Steane DA, Freeman JS, **Hudson CJ**, Myburg AA, Kilian A, Carling J, Huttner E, Sansaloni CP, Petroli CD, Grattapaglia D, McKinnon GE and Potts BM (2009). DArT markers for genomic studies in *Eucalyptus*. IUFRO Tree Biotechnology Conference. 28th June - 2nd July Whistler, BC Canada. Abstract page 8.

Technical reports

Hudson, CJ and Vaillancourt, RE (2010) Genetic structure in *Eucalyptus nitens* and ancestral identity of unknown samples using Diversity Arrays Technology (DArT) molecular markers. Confidential report prepared for Gunns Pty Ltd.

Table of contents

Preface

The genus *Eucalyptus* is a member of the *Myrtaceae* family and contains 700+ species which are mostly endemic to the continent of Australia (Williams and Brooker 1997). Eucalypt species play an important ecological role in Australia, dominating a diverse range of landscapes and habitats, ranging from wet and dry temperate forests to subalpine woodlands and tropical savannah (Byrne 2008; Eldridge *et al.* 1993). Additionally, their fast growth rates and adaptability, in combination with excellent wood and pulp qualities have seen eucalypt species and their hybrids become the most widely planted hardwood crop world-wide for commercial forestry production (Eldridge *et al.* 1993; Grattapaglia and Kirst 2008). Many of the most commercially important species belong to subgenus *Symphyomyrtus*, the largest of 13 subgenera recognised within the genus (Brooker 2000; Myburg *et al.* 2007). In particular, species and their hybrids from *Symphyomyrtus* sections *Latoangulatae* (e.g. *E. grandis* and *E. urophylla*), *Maidenaria* (e.g. *E. globulus and E. nitens*) and *Exsertaria* (e.g. *E. camaldulensis*) account for most of the 20 million hectares of eucalypt plantations established world-wide for paper, fuel-wood and solid timber production (Doughty 2000; Eldridge *et al.* 1993; Iglesias-Trabado and Wilstermann 2008; Myburg *et al.* 2007).

Eucalypts are diploid plants with a haploid chromosome number of eleven (Bachir and Abdellah 2006; Eldridge *et al.* 1993). Flow cytometry has been used to estimate the genome sizes of several *Symphyomyrtus* species, with estimates ranging between 530 Mbp (*E. globulus* and *E. dunnii*) to 710 Mbp (*E. saligna*; Grattapaglia and Bradshaw 1994). A variety of molecular marker types have been used in eucalypt studies to investigate genetic diversity within commercially important and rare and threatened species, resolve taxonomic uncertainty, and assess the level of genome similarity (*i.e.* synteny and colinearity) between species (see reviews by Byrne 2008; Grattapaglia *et al.* submitted; Myburg *et al.* 2007; Poke *et al.* 2005). The earliest molecular studies employed isozyme (reviewed in Moran 1992) and chloroplast DNA restriction fragment length polymorphism (RFLP; Steane *et al.*1991) markers to investigate genetic diversity and mating systems. Soon after, random amplified polymorphic DNA (RAPD) markers were used to construct the first linkage maps in the genus (Grattapaglia and Sederoff 1994). Marker systems have since evolved rapidly, and more recent studies have employed more sophisticated markers, such as higher throughput amplified fragment length polymorphism (AFLP; Myburg *et al.* 2003) and co-dominant microsatellite

markers (Brondani *et al*. 1998; Brondani *et al*. 2002; Brondani *et al*. 2006; Byrne *et al.* 1996; Faria *et al.* 2011; Glaubitz *et al.* 2001; Steane *et al.* 2001).

While previous studies have contributed greatly to our understanding of genome organisation and genetic variation in eucalypts, the resolution and information provided by these studies has often been limited by the molecular markers available. For example, many marker types suffer limitations due to their low level of polymorphism, transferability, and or reproducibility provided across studies (e.g. RAPDs) and or the intensive labour requirements for marker development (e.g. microsatellite development) or genotyping (Brondani *et al.* 2006; Collard *et al.* 2005; Gupta *et al.* 2008). Thus, to increase the resolution and speed of eucalypt genomics research, high throughput genotyping methods which provide wide genome coverage are needed (Sansaloni *et al.* 2010). Recently developed hybridisation-based genotyping methods in which nucleic acids are immobilised on solid-state surfaces (e.g. microarrays; Gupta *et al.* 2008; Jaccoud *et al.* 2001) provide a means to fulfil these requirements.

In this study, genomic studies in *Symphyomyrtus* species were conducted using a recently developed Diversity Arrays Technology (DArT) microarray genotyping system for use in *Eucalyptus* (Sansaloni *et al.* 2010; Steane *et al.* 2011). The initial proof of concept of DArT technology was provided using rice and published by Jaccoud *et al*. (2001). This technology has since been used extensively for linkage mapping and associated quantitative trait loci (QTL) studies in a diverse range of species; including barley, wheat, sugarcane, *Sorghum* and banana (reviewed in Gupta *et al.* 2008). The DArT array developed for *Eucalyptus* contains 7680 markers and was developed from the DNA of 64 eucalypt species (Sansaloni *et al.* 2010; Steane *et al.* 2011). It provides a high-throughput platform for the rapid genotyping of many individuals and due to the high transferability of markers across species has the potential to provide many hundreds to thousands of common polymorphic markers between studies (Sansaloni *et al.* 2010). In brief, markers on the array were developed through a complexity reduction method of total DNA using *Pst*I/*Taq*I restriction enzymes. Following digestion, the resulting DNA fragments were cloned and a prototype array was produced in order to identify informative polymorphic markers. Following this and redundancy analyses to identify and remove redundant clones (e.g. duplicates of the same fragment) an operational array was produced (Sansaloni *et al.* 2010). In sample assays, sample DNA is firstly digested using the same *Pst*I/*Taq*I restriction enzymes used in marker development. Fragments are then fluorescently labelled and hybridised to microarrays.

Following hybridisation the microarrays are scanned and marker hybridisation intensities are analysed using DArTsoft (version 7.44) software (http://www.diversityarrays.com/). Marker presence (1) or absence (0) scores and marker quality parameters are then generated for each dominant bi-allelic DArT marker (Sansaloni *et al.* 2010). Single nucleotide polymorphisms (SNPs) or indels occurring within *Pst*I/*Taq*I restriction sites, and / or indel or other rearrangements occurring within the DArT fragment itself (Jaccoud *et al.* 2001), create marker polymorphisms between individuals.

Despite the benefits provided by a high throughput and readily transferable genotyping system, DArT markers do suffer from some limitations. For example, dominant markers provide less segregation information for linkage mapping studies and the ability to dissect QTL effects is limited in comparison to more informative co-dominant markers. Furthermore, it is difficult to estimate allele frequencies which can limit their use for population genetic analyses (e.g. estimating inbreeding). However, these limitations are expected to be greatly out-weighed by the resolution provided by the DArT markers. Additionally, most of the 7680 markers contained on the genotyping array have been sequenced (GenBank accession numbers HR865291 - HR872186) and can therefore provide a direct link (through BLAST searches) to the recently released assembled *E. grandis* genome sequence (V1.0 released January 2010; www.phytozome.com/).

Chapter 1 of this thesis details linkage map construction in a large *E. globulus* dwarf x tall ecotype out-crossed F_2 family (n = 503) using DArT and microsatellite markers. This map was used in comparative mapping analyses to assess genome synteny and colinearity with *E. grandis*. In Chapter 2, seven independently constructed linkage maps were integrated into a single composite linkage map. Quantitative trait loci (QTL) analyses were conducted in Chapter 3 to investigate the underlying genetic architecture of traits which differentiate tall and dwarf ecotypes of *E. globulus*. Finally, in Chapter 4 a pilot population genomics study was conducted using five *Symphyomyrtus* species. This study aimed to assess the ability of DArT markers in combination with the composite map (Chapter 2) to identify the genomic regions and molecular basis of species differences. The final Conclusions section summarises the key results of this thesis.

The following chapter has been removed for copyright or proprietary reasons

Chapter 1 High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping

Published as:

Hudson CJ, Kullan ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM, Vaillancourt RE (2011) High synteny and colinearity among Eucalyptus genomes revealed by high-density comparative genetic mapping. Tree Genetics & Genomes (http://dx.doi.org/10.1007/s11295-011-0444-9)

Chapter 2 A composite genetic linkage map for *Eucalyptus* containing 4229 markers derived from seven independent pedigrees and 1898 individuals

2.1 Introduction

Genetic linkage mapping determines the relative positions of, and distances between, genes or DNA-based markers along linkage groups based on recombination values obtained in mapping pedigrees (Jones *et al.* 1997; Semagn *et al.* 2006). Linkage maps constructed from segregation data are valuable tools and in many cases an important precursor to many genetic analyses. For example, linkage maps provide a basis for quantitative trait loci (QTL) analyses which aim to identify molecular markers (or genes), along with their genomic position, that are linked to important phenotypic traits (Anderson *et al.* 2011; Freeman *et al.* 2008; Kearsey and Farquhar 1998). Results from QTL studies may provide information that can be subsequently be applied in marker-assisted breeding programmes or help guide the choice of candidate genes to be used in association studies (Brown *et al.* 2003; Wheeler *et al.* 2005); which may aim to address conservation and or population genetics related questions (Myburg *et al.* 2007). Linkage maps can also provide a framework for anchoring physical maps and can assist in the assembly of genome sequences (Kullan *et al.* 2010; Semagn *et al.* 2006). Furthermore, the similarity of species genome organisation can also be investigated through comparative mapping studies (Krutovsky *et al.* 2004; Lefebvre-Pautigny *et al.* 2009; Paterson *et al.* 2000).The wide application and high value of genetic linkage maps to genetics research has consequently led to numerous linkage mapping projects being undertaken in plants. For example, detailed linkage maps have been produced for all of the world's staple cereal species (Jones *et al.* 2009) and in forest trees, linkage maps have been produced for many of the most widely-planted species due to their commercial importance as wood and fibre crops (Grattapaglia *et al.* 2009; Krutovsky *et al.* 2004; Poke *et al.* 2005).

In the genus *Eucalyptus*, many pedigrees have been established for the purpose of linkage map construction and associated QTL analyses. Following the first genetic linkage map produced in the genus by Grattapaglia and Sederoff (1994), more than 20 genetic linkage maps have been published in *Eucalyptus*. Most of these have been produced in the main commercially grown species, or their hybrids, from the subgenus

Symphyomyrtus. Thus, the majority of linkage mapping projects have focussed on *E. grandis*, *E. urophylla* and *E. globulus* (reviewed in Grattapaglia *et al.* submitted) with a small number of maps having also been produced in *E. nitens* (Byrne *et al.* 1995), *E. teriticornis* (Gan *et al.* 2003; Marques *et al.* 1998), *E. camaldulensis* (Agrama and Salah 2002) and in the closely related genera *Corymbia* (Shepherd *et al.* 2006).

A variety of molecular marker types have been used in the construction of eucalypt linkage maps (Grattapaglia *et al.* submitted; Myburg *et al.* 2007). Many early maps utilised polymerase chain reaction (PCR) based markers such as random amplification of polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) molecular markers. However, the anonymous nature of these dominant marker types and the low proportion of shared polymorphism between pedigrees have limited the transfer of linkage information between studies using these markers (Brondani *et al.* 2006; Grattapaglia *et al.* submitted). More informative, codominant markers have also been used successfully in eucalypt linkage mapping. However, the low inter-pedigree polymorphism and labour intensive genotyping requirements of isozyme and random fragment length polymorphism (RFLPs) markers have limited their use for routine mapping in multiple pedigrees (Brondani *et al.* 2006; Grattapaglia *et al.* submitted). The more recent development of numerous highly polymorphic microsatellite markers made available a large suite of markers that were transferrable between species and polymorphic in multiple pedigrees. This enabled linkage group synteny to be established between maps containing common microsatellite markers and the positions and stability of QTL across multiple species to be examined (e.g. Freeman *et al.* 2009; Gion *et al.* 2011; Marques *et al.* 2002; Thumma *et al.* 2010). The ability to establish linkage group synteny has also enabled moderate-density comparative mapping studies to be performed (Brondani *et al.* 2006; Myburg *et al.* 2003). Although these linkage mapping studies have contributed substantially to eucalypt genetics research, most maps have been of relatively low density (e.g. generally comprising less than 400 markers; Grattapaglia *et al.* submitted) which has prohibited more detailed comparative analyses.

Recent advances in molecular methods have lead to high-throughput genotyping systems becoming widely used for genotyping large suites of molecular markers (Collard *et al.* 2009). Using these systems, it is possible to generate many hundreds of markers in a single mapping pedigree and this has consequently resulted in the construction of higher density linkage maps while providing more shared markers between studies (Jones *et al.* 2009). In *Eucalyptus*, Diversity Arrays Technology (DArT)

(Jaccoud *et al.* 2001) has been most recently used in genetic linkage mapping (this thesis Chapter 1; Freeman unpublished; Hudson *et al.* in press; Kullan *et al.* in press; Sansaloni *et al.* 2009). DArT is a restriction fragment, hybridisation, array-based molecular marker system which enables high-throughput genotyping combined with wide genome coverage (Sansaloni *et al.* 2010; Steane *et al.* 2011). The DArT markers are highly transferable across eucalypt species and are expected to provide many hundreds of shared markers between any two mapping pedigrees within subgenus *Symphyomyrtus* (Sansaloni *et al.* 2010; Steane *et al.* 2011). For comparative analyses between pedigrees, this greater number of shared markers provides much higher resolution and will be of particular benefit when comparing the positions of QTL detected for the same (or correlated) traits in different studies. In early QTL studies, the ability to validate the expression of QTL across variable genetic backgrounds has been hampered by a lack of shared markers (Brondani *et al.* 2006). For example, Thumma *et al.* (2010) recently detected multiple co-locating growth-related QTL on linkage group 5 in an *E. nitens* family but was unable to accurately compare the position of this QTL to similar growth-related trait QTL detected on this same linkage group in two other studies (Freeman *et al.* 2009; Grattapaglia *et al.* 1996) due to a lack of shared markers. The use of DArT markers for linkage mapping offers the potential to overcome these limitations, by providing a large number of common markers between studies and enabling comparative analyses to be performed on an unprecedented scale. Another benefit of the DArT markers is that most of the 7680 markers contained on the genotyping array have been sequenced and have been made publicly available (GenBank accession numbers HR865291 - HR872186). Therefore, it will be possible to identify a DArT marker positions in the *E. grandis* genome sequence (V1.0 released January 2011; www.phytozome.net/) and this information, for example, may assist in the identification of positional candidate genes.

To date, DArT marker linkage maps have been constructed for *E. globulus* and *E. grandis* x *E. urophylla* hybrid families in seven independent mapping pedigrees (this thesis Chapter 1; Freeman unpublished; Hudson *et al.* in press; Kullan *et al.* in press; Sansaloni *et al.* 2009). In the two largest of these mapping families (more than 500 individuals each), 1010 (this thesis Chapter 1; Hudson *et al.* in press) and 2229 (Kullan *et al.* in press) DArT markers, respectively, have been mapped at sub-centimorgan marker densities. While the use of DArT markers in these mapping families have facilitated the construction of high density linkage maps, DArT markers do suffer some

limitations due to their dominant nature. For example, the ability to construct sex-averaged linkage maps (*i.e.* consensus within a pedigree) and dissect QTL effects with DArT markers is limited in comparison to more informative co-dominant markers. Additionally, dominant markers are less useful for some other genetic analyses, such as estimating population genetic parameters (e.g. inbreeding levels). Due to these limitations, a variable number of co-dominant microsatellite markers have also been mapped in all of the linkage maps constructed with DArT. An additional benefit of mapping microsatellite markers in these maps is that these markers provide a link to many of the earlier non-DArT marker linkage maps produced in *Eucalyptus* which mapped these markers. A total of more than 4000 DArT and microsatellite markers have been collectively mapped in the seven eucalypt pedigrees described above. To enable the relative positions of markers and QTL from different studies to be determined more easily, it would be of great benefit to integrate all of this mapping data into a single map. This would also provide a valuable reference map for *Symphyomyrtus* and related species.

All of the eucalypt linkage maps constructed using DArT markers were produced using the program JoinMap 4.0 (Van Ooijen 2006); which is one of the most commonly used programs for constructing linkage maps in plant populations. This program also appears to be the only software available for building linkage maps using the combined segregation data from multiple populations (Cheema and Dicks 2009; Li *et al.* 2010; Wenzl *et al.* 2006). However, it is presently not feasible to combine the segregation data contained within the seven eucalypt mapping families describe above (collectively containing 1898 individuals), and successfully order such large numbers of markers within linkage groups (up to ~ 500) using JoinMap due to computational limitations (Van Ooijen *pers comm.*). In this situation, this task is made even more difficult because the majority of markers in these datasets are dominant marker-types (e.g. DArT markers), which have incomplete segregation information and result in the exponential increase of marker-ordering calculations compared to fully-informative co-dominant markers (Van Ooijen 2006). In contrast to traditional segregation-data based methods of linkage map construction (e.g. JoinMap), alternative strategies have been developed for multi-pedigree map construction in which markers are simply merged onto linkage groups using the marker position information of mapped markers relative to other markers. For example, the 'neighbours' marker-merging approach of Cone *et al*. (2002) and the marker-merging method implemented in the PhenoMap program (GeneFlow Inc.

USA) have been used to successfully construct high density linkage maps using multiple mapping pedigree linkage map data (containing many several thousand markers) in a number of plant species; including *Sorghum* (Mace *et al.* 2009), barley (Alsop *et al.* 2011; Varshney *et al.* 2007; Wenzl *et al.* 2006) and maize (Cone *et al.* 2002; Wei *et al.* 2007).

In this study, a marker-merging method is used to construct a high-density DArT and microsatellite map using marker position information from seven independent eucalypt pedigrees. Recent comparative mapping analyses using 236 to 393 shared markers between three of the linkage maps used to construct this map (*E. urophylla* x *E. grandis* SA double pseudo-backcross F_2, *E. globulus* Lighthouse F_2 and *E. globulus* FAM1 F_2 mapping families; this thesis Chapter 1 and Hudson *et al.* in press) showed that these linkage maps exhibited both high synteny (> 93.4% markers occurring on the same linkage groups) and high colinearity (> 93.7% markers having the same order within linkage groups). This suggests that it should be possible to merge map information from several component maps into a single high quality map featuring robust marker-order together with very high marker density.

2.2 Materials and Methods

As the terminology used to describe different types of linkage maps is inconsistent throughout the linkage mapping literature, the following terms used in this Chapter are defined as; (1) sex-averaged map – a consensus map of individually constructed male and female maps from a single family, (2) consensus map – a consensus map that integrates multiple (*i.e.* from more than one progeny), independently constructed, male and female maps (e.g. F_2 double-pseudo backcross), and (3) composite map – a map produced through the integration of multiple independently constructed sex-averaged and or consensus maps.

The *Eucalyptus* multi-species composite map (hereafter referred to as the 'composite map') was built using map data from an *E. grandis* x *E. urophylla* F_2 double pseudo-backcross pedigree consensus map (Kullan *et al.* in press) plus one *E. urophylla* x *E. grandis* (Sansaloni and Petroli unpublished) and five *E. globulus* (this thesis Chapter 1; Freeman unpublished; Hudson *et al.* in press) sex-averaged linkage maps which were constructed in outcrossed F_2 families (hereafter referred to as 'component' maps). Component map family sizes ranged from 121 (GLOB-F_2-1) to 547 (GU-SA) and collectively contained 1,898 individuals (Table 2.1). Component maps were constructed by different researchers. All used the program JoinMap 4.0 (Van Ooijen 2006) with marker-ordering within linkage groups (LGs) estimated using the regression algorithm of Stam (1993) combined with the Kosambi mapping function. All component maps comprised 11 LGs in accordance with the haploid chromosome number of *Eucalyptus* and had similar total map lengths (1033-1258 cM; Table 2.1 and supplementary CD File 1 available on compact disc at the end of this thesis). The numbering and orientation of composite map LGs followed Brondani *et al.* (2006); this also corresponds to the assembled *E. grandis* genome sequence available at www.phytozome.net/ (V1.0 released January 2011).

Table 2.1 The main features of the component maps used to construct the *Eucalyptus* multi-species composite map, including; mapping pedigree progeny size (*n*), map length (cM), mean marker interval (MMI) and total number of mapped markers[d]. For DArT, microsatellite (SSR; short sequence repeat) and gene markers mapped on each component map, the percentage of markers unique to that map (*i.e.* not mapped in any of the six other component maps) is shown in parentheses.

Name and reference[a]	Map abbreviation	*n*	cM[b]	MMI[c]	Markers mapped[d] (percentage of unique markers in pedigree)			
					DArT	SSR	Gene	Total
E. urophylla x *E. grandis* SA double pseudo-backcross F_2[e]	GU-SA	547	1107	0.51	2229 (45%)	59 (46%)	2 (100%)	2290 (45%)
E. urophylla x *E. grandis* Embrapa F_2[fi]	GU-Emb	177	1229	0.78	1617 (41%)	193 (77%)	0	1810 (44%)
E. globulus Lighthouse F_2[g]	GLOB-LH	503	1151	1.21	1010 (27%)	50 (12%)	0	1060 (27%)
E. globulus FAM1 F_1[h]	GLOB-F_1-1	183	1033	1.97	571 (14%)	4 (0%)	2 (0%)	577 (14%)
E. globulus FAM4 F_1[h]	GLOB-F_1-4	183	1137	2.46	488 (10%)	6 (0%)	4 (25%)	498 (10%)
E. globulus FAM5 F_1[h]	GLOB-F_1-5	184	1055	2.09	600 (22%)	4 (0%)	2 (0%)	606 (21%)
E. globulus FAM1 F_2[h]	GLOB-F_2-1	121	1258	2.73	660 (18%)	30 (30%)	5 (40%)	695 (18%)

[a]Includes species, cross type and reference; [e]Kullan *et al.* (in press) ; [f]Sansaloni and Petroli (unpublished); [g]this thesis Chapter 1 and Hudson *et al.* (in press); [h]Freeman (unpublished). [b]Map length; total for all 11 LGs. [c]MMI; mean marker interval between adjacent loci, average for all 11 LGs. [d]Total number of component map markers calculated using only those LGs that were included in construction of composite maps; see following text in materials and methods and Table 2.2. [i]Data for the *E. urophylla* x *E. grandis* Embrapa F_2 component map calculated using a combination of framework and comprehensive LGs; see following text in materials and methods and Table 2.2.

Prior to building the composite map, marker names were standardised and the colinearity of common markers between homologous LGs of component maps were visually inspected by plotting map data in the program MapChart (Voorrips 2002). Five LGs from three component maps were found to have substantial regions of non-colinearity (discordant marker-orders) with other component maps. These five LGs from three of the smaller *E. globulus* mapping families (Table 2.1) were excluded from composite map construction in order to maximise marker-order accuracy (Table 2.2). In the *E. urophylla* x *E. grandis* Embrapa F_2 linkage mapping family (GU-Emb), a framework mapping strategy (Keats *et al.* 1991) was used resulting in two maps, a 1032-marker framework and a 2484-marker comprehensive map. A number of GU-Emb comprehensive map LGs exhibited substantial non-colinearity with other component maps, and in some cases, with the corresponding framework LG. In order to maximise both the total number of markers placed, as well as agreement in marker-order between maps, a combination of GU-Emb framework and comprehensive LGs were used for composite map construction (Table 2.2).

Table 2.2 Details of component map LGs included in the construction of the *Eucalyptus* multi-species composite map. All 11 LGs of the *E. urophylla* x *E. grandis* Embrapa F_2 pedigree were used; letters indicate whether framework (F) or comprehensive (C) LGs were used (see preceding text in materials and methods).

Map	Linkage group											
	1	2	3	4	5	6	7	8	9	10	11	Total
E. urophylla x *E. grandis* SA double pseudo-backcross F_2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11
E. urophylla x *E. grandis* Embrapa F_2	F	C	F	C	F	C	F	C	F	C	C	11
E. globulus Lighthouse F_2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11
E. globulus FAM1 F_1	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	9
E. globulus FAM4 F_1	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	9
E. globulus FAM5 F_1	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	10
E. globulus FAM1 F_2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11
Total	7	7	7	5	7	6	7	7	6	6	7	

For component map LGs included in composite map construction (Table 2.2), the number of markers ranged from 498 (GLOB-F_1-4) to 2290 (GU-SA; Table 2.1). In total, component maps contained 4350 individual markers, which included 4089 DArT, 253 microsatellites (6 CRC, 35 CSIRO and 212 Embrapa) and 8 mapped genes. Ninety-six markers (2.2%) were mapped to two or more linkage groups (*i.e.* 'multicopy' markers)

across component maps. This resulted in the 4350 individual markers being mapped to 4457 locus positions. Of these 4457 positions, 1960 could be considered to be bridging loci, meaning that markers at these loci had been mapped to syntenic linkage groups in two or more component maps and would serve as anchor points between maps during composite map construction. Conversely, 2497 marker positions were unique to a single component map.

The composite linkage map was constructed at Diversity Array Technology (DArT) Pty Ltd (Canberra, Australia) using specially developed R scripts which merged component map markers into the composite map based on their relative mapped position. The *E. urophylla* x *E. grandis* SA F_2 (GU-SA) linkage map was used as the (base) seed-map in composite map construction due to it having the largest progeny size, the largest number of both mapped and unique markers (Table 2.1) and due to it being used to order contigs along chromosomes in the *E. grandis* sequencing project (Kullan *et al.* in press). In map construction, the seed-map acted as a 'fixed backbone' to which markers from other component maps were added. The procedure for building the composite map was as follows. The number of common markers between homologous seed-map and component map LGs were calculated. Spearman rank marker-order correlations were then calculated and a component map 'fit value' was calculated for each linkage group as; Fit value = correlation x log (number of common markers). The component map LG having the highest fit value was then selected and unique markers from this LG (*i.e.* those not mapped on the seed map, or the 'building' composite map) were added to the seed-map LG (or building composite map) using linear regression models. Once markers had been added to the seed-map in this so-called first round, homologous LGs of the remaining component maps were then compared to this new map (*i.e.* the 'building' composite) and the process was repeated. This continued until all unique markers had been added from remaining component map LGs which met the following criteria; (1) shared at least three common markers with the homologous LG of the seed-map, and (2) the marker-order correlation coefficient exceeded 0.50.

By constructing the composite map in this manner, it was possible for some or all markers of a single component map to be included in a composite map LG without using, or incorporating, the 'marker-order' information from that component map. This could occur when all markers of a component map were added to the composite in earlier building rounds; e.g. markers were mapped on the seed-map or were added from other component maps having better fit values.

A second composite map was also built using the *E. globulus* Lighthouse F_2 (GLOB-LH) linkage map as the seed-map. For each LG, marker-order correlations between composite and component maps were calculated in SAS 9.2 (SAS Institute, Cary, USA) using the PROC CORR Spearman function. The distribution of multicopy marker positions over all LGs (homogeneity) was tested by a chi-square test; by comparing the number of multicopy marker positions occurring on each linkage group (observed frequency) against an expected frequency calculated for each LG as; (total number of multicopy marker positions in the composite map / total number of DArT marker positions in the composite map) x number of DArT marker positions per LG. Evidence for DArT multicopy marker loci duplications in the *E. grandis* genome sequence were investigated using the BLAST server available at www.phytozome.net/. The BLAST (bl2seq) tool available at http://www.ncbi.nlm.nih.gov/ was used to investigate DArT sequence similarity/marker redundancy. All graphical representations of linkage maps were drawn using MapChart.

2.3 Results

The composite map contained 4135 individual markers, comprising 3909 DArT markers, 8 genes and 218 microsatellites (177 Embrapa, 6 CRC and 35 CSIRO microsatellites). The total map length of the composite was 1237 cM and was within the range of component map lengths (1033-1258 cM; Table 2.1). Although nine out of eleven LGs contained a small number of marker intervals greater than 5 cM (range 1-5; Table 2.3) the average marker interval across all 11 LGs was 0.32 cM (Table 2.3). Over half of the markers mapped in the composite map (2269 or 55%) had been mapped in a single component map only and 84 DArT markers (2.08% of total DArT markers) mapped to multiple LGs (multicopy markers; Table S2.1). Most multicopy markers had been mapped to two individual LGs, however two (ePt-574238, ePt-643458) and four DArT markers (ePt-503174, ePt-568818, ePt-637610, ePt-637861) were mapped to three and four LGs, respectively (supplementary Table S2.1). This resulted in the 4135 markers being mapped to 4229 positions (Table 2.3). The number of multicopy DArT marker positions on each LG ranged from 5 to 24, which represented 2-6.5% of the total number of DArT markers mapped per LG (Table 2.3). Although LGs five and seven contained a larger proportion of multicopy DArT marker positions (e.g. LG1 contained only 5 multicopy DArT marker positions, or 2% of the total number of DArT marker positions; Table 2.3), the proportion of multicopy DArT marker positions found on each LG did not significantly differ from that expected by chance across all LGs ($\chi^2 = 12.47$, $P = 0.25$, $df = 10$). There was no trend within LGs for multicopy DArT markers to be clumped in either distal or central LG areas (data not shown). Details of the markers mapped in the composite map, including the mapped position of all loci, their component map(s) origin and details of multicopy DArT markers are presented in supplementary CD File 2.

Table 2.3 Summary of the *Eucalyptus* multi-species composite map, showing LG length (cM), number of mapped markers per LG, and marker interval (cM) information (Av. = average). MC DArT pos. indicates the number of multicopy (MC) DArT marker positions (pos.) occurring on each LG and the percentage (%) of total DArT markers per LG; *i.e.* 5 multicopy DArT marker positions occurred on LG1, being 2% of the total number of DArT markers on this LG.

LG[a]	cM	Markers mapped				Marker interval distance (cM)[d]				MC DArT
		DArT	SSR	Genes	Total	Av.	1< 2	2 < 5	≥ 5	pos. (%)[e]
1[b]	99.3	251	12	0	263	0.41	16	7	3	5 (2.0)
2[b]	102.1	451	29	0	480	0.22	11	2	0	19 (4.2)
3[b]	111.1	430	18	2	450	0.27	19	1	1	22 (5.1)
4[c]	79.0	231	9	3	243	0.36	7	7	0	12 (5.2)
5[b]	110.5	368	8	0	376	0.32	12	6	3	24 (6.5)
6[b]	136.9	410	43	1	454	0.33	12	6	2	15 (3.7)
7[b]	97.1	306	9	1	316	0.33	17	4	1	19 (6.2)
8[b]	137.3	540	26	0	566	0.26	15	4	2	19 (3.5)
9[b]	93.9	314	20	0	334	0.31	16	3	2	12 (3.8)
10[b]	142.9	347	20	1	368	0.42	14	6	5	13 (3.7)
11[b]	127.0	355	24	0	379	0.37	20	2	2	18 (5.1)
Total	*1237.1*	*4003*	*218*	*8*	*4229*	*0.32*	*159*	*48*	*21*	*178*[f]

[a]Linkage group (LG) of the *Eucalyptus* multi-species composite map; [b]GU-SA or [c]GLOB-LH seed-map used for composite map construction. [d]Maker interval distance; average maker interval per LG and number of marker intervals between 1 < 2 cM, 2 < 5 cM and ≥ 5 cM. [e]MC DArT pos. shows the number of multicopy (MC) DArT marker positions (pos.) occurring on each linkage group. The percentage (%) is calculated as MC DArT pos. / Total DArT positions*100 per LG. [f]The 178 multicopy DArT marker positions represent 84 multicopy DArT markers (see Table S2.1).

LG colinearity between component maps and the composite map can be seen graphically in supplementary Figure S2.1A-L. Pair-wise LG marker-order correlations were generally high (greater than 0.90; Table 2.4) and was reflected by the high colinearity shown between common markers in Figure S2.1A-L. However, a small degree of non-colinearity did occur between all component maps and the composite map. Eleven component map LGs had marker-order correlations of less than 0.90 (Table 2.4), however, these LGs were either, (1) identified as having poor marker colinearity with other component maps prior to composite map construction and excluded from analysis (gray shading Table 2.2 and Table 2.4), or (2) not incorporated during construction of the multi-species composite map (correlation value without asterisk; Table 2.4). Thus, these poorly correlated LGs did not adversely affect the composite map marker-order. For each LG, the average pair-wise marker-order correlation between the composite map and those component maps included in map construction ranged from 0.91 (LG1 and LG4) to 0.98 (LG11, Average[b] column; Table 2.4).

Table 2.4 Linkage group (LG) pair-wise marker-order correlation coefficients between component maps and the *Eucalyptus* multi-species composite map. For comparison, LG4 correlations are shown for both GU-SA and GLOB-LH seed-map composite maps. For the GU-Emb component map, superscript letters indicates whether the framework (F) or comprehensive (C) LG was used in map construction. Gray shading indicates the component map LGs that were excluded from composite map construction. An asterisk following the correlation value indicates that marker-order information from this component map was incorporated during construction of the composite map LG. Apart from two correlations (indicated by ^{ns} superscript) all correlations were significant at $\alpha \leq 0.05$.

LG	Composite LG seed-map used	Component map								
		GLOB-LH	GU-Emb	GLOB-F_2-1	GLOB-F_1-1	GLOB-F_1-4	GLOB-F_1-5	GU-SA	Average[a]	Average[b]
1	GU-SA	0.978*	0.560^{F}	0.989*	0.994*	0.980*	0.952*	1.000*	0.909	0.979
2	GU-SA	0.981*	0.953*^{C}	0.929*	0.952*	0.983*	0.848	1.000*	0.941	0.957
3	GU-SA	0.908*	0.989*^{F}	0.973*	0.991*	0.969*	0.988*	1.000*	0.970	0.982
4	GU-SA	0.957*	0.875*	0.849*	0.827	0.802	0.182^{ns}	1.000*	0.749	0.877
4	GLOB-LH	1.000*	0.744^{C}	0.951*	0.943*	0.834	0.224^{ns}	0.997*	0.783	0.909
5	GU-SA	0.995*	0.918^{F}	0.959*	0.960*	0.988*	0.959*	1.000*	0.963	0.972
6	GU-SA	0.991*	0.994*^{C}	0.990*	0.623	0.988*	0.861	1.000*	0.908	0.965
7	GU-SA	0.984*	0.652^{F}	0.988*	0.974*	0.962*	0.910*	1.000*	0.912	0.964
8	GU-SA	0.984*	0.990*^{C}	0.992*	0.656	0.945*	0.991*	1.000*	0.926	0.980
9	GU-SA	0.994*	0.972*^{F}	0.978*	0.972*	0.650	0.968	1.000*	0.922	0.977
10	GU-SA	0.951*	0.984*^{C}	0.972*	0.354^{ns}	0.925	0.971*	1.000*	0.860	0.961
11	GU-SA	0.988*	0.992*^{C}	0.991*	0.974*	0.957	0.993*	1.000*	0.983	0.988
Average[c]		0.978	0.886	0.974	0.854	0.926	0.879	1.000		
Average[d]		0.983	0.982	0.974	0.935	0.966	0.944	1.000		

Pair-wise averages: [a]calculated using all six component maps (composite LG seed-map excluded), [b]calculated using only those LGs included in composite map construction (marked with an asterisk, seed-map excluded), [c]calculated using all 11 LGs, [d]calculated using only those LGs included in composite map construction (marked with an asterisk, seed-map excluded).

The composite map comprised LGs 1-3 and 5-11 from the GU-SA seed-map composite map and LG4 from the GLOB-LH seed-map composite map. Although the GU-SA linkage map was the preferred seed-map (see methods and materials), LG4 of the GLOB-LH seed-map composite map provided a better representation of component map marker-orders and was considered to be of a higher quality relative to the GU-SA seed-map. A similar number of markers were mapped in both composite maps and overall their marker-orders were very similar (see supplementary Figure S2.2A-C). For example, the GU-SA seed-map composite map contained 4167 markers mapped to 4264 positions and the GLOB-LH seed-map composite map contained 4153 markers mapped to 4248 positions. However, the total map length of the GLOB-LH composite map was highly inflated (1517 cM) in comparison to the GU-SA seed-map composite map (1238 cM). The two maps shared 4114 common markers (representing 4211 positions) and the pair-wise LG marker-order correlations were mostly very high; ranging between 0.985-0.999 (all $P < 0.0005$) for LGs 1-3 and 5-11. However, the marker-order correlation for LG4 was substantially lower in comparison to other linkage groups (0.965, $P < 0.0005$; calculated on 231 common markers). This was primarily caused by a large region (~30 cM) of non-colinearity at the top of this LG (Figure S2.2A).

The poor correlation of marker-order between composite maps for LG4 appeared to result from the inclusion of marker-order information from the GU-Emb component map during construction of the GU-SA composite map. During the construction of LG4 using the GU-SA seed-map, the first markers added to this map were from the GU-Emb component map; despite these maps having a relatively poor marker-order correlation of 0.775 (data not shown). In total, markers from three component maps were added to LG4 of the GU-SA seed-map during composite map construction (see Table 2.4). The average pair-wise correlation calculated for these three maps with the 'building' composite LG was 0.803 (range 0.775 to 0.848; data not shown). In contrast, the average pair-wise marker-order correlation of the three maps included in construction of LG4 using the GLOB-LH seed-map composite map was considerably higher at 0.919 (range 0.862 to 0.969; data not shown) similarly calculated using the 'building' composite map. Marker-order information from the GU-Emb map was not incorporated in LG4 of the GLOB-LH seed-map composite map.

Pair-wise marker-order correlations calculated using the 'constructed' LG4 of both seed-map composite maps were higher for four out of five component maps (seed-map component maps excluded) with the GLOB-LH seed-map composite map relative to the

GU-SA seed-map composite map (Table 2.4). Thus, the marker-order of the GLOB-LH seed-map composite map more accurately represented the component map marker-orders for this LG (see supplementary Figure S2.1D and Figure S2.1E). Although LG4 of the GLOB-LH seed-map composite map (243 markers) contained 35 less markers than the GU-SA seed-map composite map (278 markers), both LGs had similar lengths (both 79 cM), and due to the higher quality of LG4 in the GLOB-LH seed-map composite map, this LG was included in the composite linkage map.

Although not a main focus of this study, evidence for the occurrence of duplicated DArT marker loci within the assembled *E. grandis* genome sequence (V1.0 available at www.phytozome.net/) was investigate for the six multicopy markers which had been mapped to three or more LGs (Table S2.1). Two of these DArT markers had been mapped to identical positions in four LGs (ePt-637610 and ePt-637861; Table S2.1) and were found to be redundant markers (*i.e.* identical clones or sequences) based on their marker sequence similarity (bl2seq: 583/606 base-pair similarity, e-value: 0.0). For the five unique multicopy markers, three were detected to have loci duplications within the *E. grandis* genome sequence. In each case, the duplicated loci *E. grandis* chromosome assembly corresponded to the linkage group to which the marker was mapped.

2.4 Discussion

The integration of linkage map data from seven independent studies resulted in a composite map containing 4029 DArT, 218 microsatellite markers and 8 mapped genes. This map successfully incorporated a large amount of the genetic diversity present between the two *E. grandis* x *E. urophylla* hybrid and five pure-species *E. globulus* pedigrees used in its construction, evident by the fact that over half (55%) of the markers in the composite map were mapped in a single component map only. The composite map had a total length of 1237 cM and an average marker spacing of 0.32 cM representing the highest density map produced in the genus to date. This map will serve as a valuable reference map for *Eucalyptus* and the information contained within it offers the potential to perform more detailed and faster investigations of genome differentiation among eucalypt species, validate QTL across variable genetic backgrounds, and position a growing number of candidate genes co-localized with QTL.

A major advantage of the marker-merger method used in this study was the substantial time and labour savings made when compared to the effort required to produce comparable maps using traditional, segregation-based mapping methods. For example, Li *et al.* (2010) constructed a 2111 marker composite map from four barley mapping pedigrees using JoinMap 3.0 (Van Ooijen and Voorrips 2001) and reported that it took '*several thousand hours*' of computing time. In a larger barley study, Wenzl *et al.* (2006) produced a 2935 loci composite map from ten mapping populations using JoinMap 3.0 in combination with specially built Perl scripts to overcome some of the computer programs limitations. Similarly, this was a time and labour intensive project and required several months of semi-manual data processing (Wenzl *et al.* 2006). In contrast, the two composite maps produced in this study using a marker-merging method were built in a single day.

Similar marker-merging methods to that used in this study have been used to produce composite linkage maps in various plant species (Alsop *et al.* 2011; Mace *et al.* 2009; Varshney *et al.* 2007; Wenzl *et al.* 2006). All of these past studies have reported good agreement between composite map and individual component map marker-orders. However, only one study (Wenzl *et al.* 2006) has used the same data to produce both marker-merging and segregation-based composite maps to enable a direct comparison between marker-orders produced using both methods. Comparison of the two composite maps constructed in Wenzl *et al.* (2006), one using PhenoMap (marker-merging) and

the other using JoinMap 3.0/Perl scripts (segregation-based), revealed that both were highly similar in terms of their overall marker-order and map length. Although a different marker-merging procedure was used in this study, the comparison made by Wenzl *et al.* (2006) illustrated that high quality linkage maps can be successfully produced using marker-merging techniques.

As with all linkage mapping studies, it is important to consider the quality and marker-order accuracy of the map produced and any potential limitations that it may have. The marker-order accuracy of the composite map was assessed by comparison against individual component maps. As the composite map marker-order reflects the overall level of colinearity between each of the component maps used in the construction of each linkage group, the inclusion, or non-inclusion of a component map linkage group during map construction will influence the level of colinearity (e.g. LG marker-order correlations) observed. Additionally, factors specific to each component map may also influence the colinearity shown between each component map and the composite map. For example, the pedigree type and size (Ferreira *et al.* 2006; Hackett and Broadfoot 2003), the quality of marker data (e.g. free from genotyping errors; Hackett and Broadfoot 2003) and the number and distribution of fully informative co-dominant markers (e.g. microsatellites) in each mapping family (Collard *et al.* 2009; Ferreira *et al.* 2006), plus the stringency of marker-ordering fit statistics applied by different researchers could have influenced the accuracy of marker-ordering within component map linkage groups. Although similar mapping parameters were applied by different researchers during component map construction, the different species and pedigrees used in addition to variation in pedigree sizes and types could potentially result in localised marker-order differences. Despite this, the marker-order agreement between component maps and the composite map were mostly good and were reflected by the high pair-wise LG marker-order correlations. For example, 48 out of the 66 pair-wise component map LGs comparisons across all component maps had correlations greater than 0.950 (seed-map component map excluded; Table 2.4). However, despite these high correlations most component maps did exhibit some marker-order inconsistencies with the composite map.

A number of (mostly) single-marker marker-order inconsistencies did occur over large distances, although most marker-order disagreements were commonly found to occur among tightly grouped markers in regions of less than 5 cM. Although it is possible that these marker-order differences could be real and represent chromosomal rearrangements

between the different mapping pedigrees and or species, they are more likely to reflect marker-order inaccuracies within any of the component maps or simply be artefacts of the statistical uncertainty associated with ordering tightly linked markers (see Collard *et al.* 2009). A number of component map LGs were found to have poor marker-order correlations with the composite map. However, as these LGs were either identified as having poor marker colinearity with other component maps in the initial inspection of map data, and / or, were excluded during composite map construction, marker-order information from these LGs did not influence the marker-order of the composite map. The exclusion of these LGs and the generally high marker-order correlations observed for other pair-wise LG comparisons suggests that the composite map is of a sufficient quality to prudently facilitate the transfer of genetic information between studies which share common molecular markers.

Although the majority of markers contained on the composite map are DArT markers, which to date have only been mapped in the pedigrees included in this study, the composite map does contain many microsatellite markers (218) which can be used to establish synteny and colinearity with previously produced maps (e.g. 13 out of 22 earlier studies have mapped a variable number of microsatellites; see Grattapaglia *et al.* submitted). This may enable researchers to utilise the marker position information of the composite map and possibly use the sequence information of DArT markers to advance their research. For example, if a DArT marker is found to reside in a map region of interest (e.g. within a QTL confidence interval) it may be possible to convert the DArT marker into a single-marker (e.g. PCR-based) assay for future research purposes. Furthermore, the positions of DArT markers can be identified within the *E. grandis* genome sequence in order to take advantage of this sequence information and the various datasets which are currently being uploaded and annotated on this resource [e.g. information from large expressed sequence tagged (EST) projects]. However, when using the *E. grandis* genome sequence it is important to note that DArT marker loci duplications do exist within, and or between, pedigrees and or species. Evidence for a small number of DArT marker loci duplications in the genomes of different eucalypt pedigrees was reported in Hudson *et al.* (in press and this thesis Chapter 1) and in this study 84 multicopy DArT markers were mapped to multiple LGs in the composite map. For the five unique multicopy DArT markers investigated (those mapped to three or more LGs; Table S2.1), three were found to have duplicated loci in the *E. grandis* genome. Therefore, the occurrence of DArT marker loci duplications should be

considered when utilising marker information from the composite map and or in combination with the *E. grandis* genome sequence.

Another minor consideration for users of this composite map is that there is expected to be some degree of marker redundancy among the 4029 mapped DArT markers. This marker redundancy is a consequence of the processes used to generate the DArT markers contained on the genotyping array (Wenzl *et al.* 2006). As DArT markers were originally generated by cloning random DNA fragments from genomic representations (Sansaloni *et al.* 2009), this can result in the same genomic fragment being represented more than once on the genotyping array. As identical clones are expected to produce identical genotype scores they should map to identical (or near identical) map positions; as found for the two redundant markers (ePt-637610 and ePt-637861) identified in this study.

In conclusion, the integration of seven individual genetic linkage maps into a single composite map resulted in the construction of a DArT and microsatellite marker reference map for *Eucalyptus* subgenus *Symphyomyrtus* species. Although some small marker-order inconsistencies exist between component maps and the composite map, there is a relatively high agreement of marker-order between most LGs which indicates that the composite map represents a good estimation of the combined component map marker positions in most cases. However, at finer scales (sub-cM) marker-orders may differ between component and composite maps. Furthermore, the few substantial marker-order discrepancies observed between LGs highlights the importance for users of this composite map to firstly assess the level of colinearity between the composite map and their map. Overall, the genome coverage and marker density of the composite map greatly exceeded that achieved in any of the single mapping populations and it is expected that this composite map will provide a valuable reference map for the world-wide *Eucalyptus* research community, and more easily enable the transfer of genetic information between different studies and allowing for the integration of DArT marker information with other genomic resources.

Chapter 3 A vegetative phase-change micro-RNA underlies a major QTL which differentiates dwarf and tall ecotypes of the forest tree *Eucalyptus globulus*

3.1 Introduction

Widespread plant species are often exposed to differential selection pressures from variable biotic and abiotic factors in different parts of their range. Depending on the strength of selection and other variables, including the level of gene flow between adjacent populations, the genetic control of adaptive traits (Linhart and Grant 1996) and effective population size (Leinonen *et al.* 2009), localised adaptation may lead to the formation of intraspecific ecotypes. Ecotypic differentiation is common within plant species, resulting in populations which differ genetically in some morphological, physiological or life history trait in comparison to other populations (Daehler *et al.* 1999; Hufford and Mazer 2003; Leinonen *et al.* 2009). In forest trees, common-garden reciprocal-transplant trials containing individuals from populations throughout the range of a species have illustrated the occurrence of ecotypic differentiation in many species (reviewed by Savolainen *et al.* 2007). Generally, populations from the extremes of a species range are highly differentiated and adapted to their local environment. However, ecotypic differentiation can also occur over much smaller geographical scales, especially when populations are subject to strong selection pressures. For example, Sambatti and Rice (2006) showed that selection pressures associated with contrasting serpentine and riparian environments were sufficient to maintain highly differentiated and site-adapted populations of sunflower (*Helianthus exilus*) ~5 km apart. At even smaller geographical scales, highly differentiated populations of dwarf and normal-type cordgrass (*Spartina alterniflora*) ecotypes have been found growing within a few meters of each other (Daehler *et al.* 1999).

As found in many widespread forest trees, *Eucalyptus globulus* spp. *globulus* (hereafter referred to as *E. globulus*) displays adaptive population differentiation in numerous traits throughout its range (Dutkowski and Potts 1999; Jordan *et al.* 1994; Jordan *et al.* 1999; Jordan *et al.* 2000). *E .globulus* is commonly distributed throughout eastern and southern coastal areas of Tasmania, on Bass Strait Islands and extends into southern Victoria on continental Australia (Brooker and Kleinig 1999). It is generally found as a medium-sized to very tall (15-60 m) forest tree and based on the quantitative genetic

affinities of progenies grown from range-wide populations in field trials the species has been partitioned into 13 geographical races (Dutkowski and Potts 1999). Three atypical reproductively mature 'dwarf' populations of *E. globulus* occur in geographically disjunct locations in southeast Tasmania and at Wilsons Promontory in southeast Victoria. These populations occur in similar coastal environments on exposed granite headlands in which trees grow to an average height of less than four meters and often have a multi-stem habit (Foster *et al.* 2007). It is suggested that this dwarf phenotype is an adaptation to salt and high wind exposure, and/or, water stress in these environments (Chalmers 1992; Jordan *et al.* 2000). In addition to the significant tree form and height differences between dwarf and tall ecotypes, individuals from these populations also differ in a number of potentially adaptive traits, including in their timing to vegetative and reproductive phase-change (Jordan *et al.* 2000).

Many woody plant species are heteroblastic, meaning that they have distinct juvenile and adult morphology (Poethig 2003; Zotz *et al.* 2011). *Eucalyptus globulus* is one of the best known and studied examples of a heteroblastic plant (Hamilton *et al.* 2011) and during ontology individuals undergo a striking morphological change between juvenile and adult foliage. For example, juvenile *E. globulus* leaves are sessile and occur in opposite pairs on flanged stems which are squarish in cross-section. They also have an epicuticular wax (glaucousness) which gives juvenile leaves a bluish-green appearance. The juvenile vegetative state may persist for several years in *E. globulus* individuals (Jordan *et al.* 2000) before heteroblasty (or vegetative phase-change) occurs and highly differentiated non-glaucous, pendulous adult vegetation is produced (Brooker and Kleinig 1999). Associated with this change in morphology are many important physiological changes, including changes in leaf chemistry (Li *et al.* 1996), resistance to pests and disease (Dungey *et al.* 1997; Farrow *et al.* 1994; Steinbauer 2002) and change in leaf cellular organisation and photosynthetic capacity (James and Bell 2000; James and Bell 2001; Velikova *et al.* 2008). Intraspecific variation in the timing of vegetative phase-change has been observed in *E. globulus* and also in other woody tree species (Climent *et al.* 2006; Dutkowski and Potts 1999; Dutkowski *et al.* 2001; Jaya *et al.* 2010a), and due to the significant morphological and physiological changes associated with this developmental trait it is proposed that the timing of this transition is of adaptive value in *E. globulus* (Jordan *et al.* 2000).

The onset of reproductive growth (flowering) represents a second developmental phase-change transition in sexually reproducing plants (Bäurle and Dean 2006; Wang *et al.*

2011). The timing of this transition has significant consequences for reproductive success and is regulated by multiple environmental and internal developmental cues. This ensures that sufficient internal resources have been accumulated and that environmental conditions are favourable for flowering (Henderson and Dean 2004; Simpson *et al.* 1999). In perennial species such as *E. globulus*, flowering has two separate components, precocity (age to first flowering), and flowering time period (*i.e.* month(s) of flowering). The age at which eucalypt species first produce flowers varies markedly both between and within species (Eldridge *et al.* 1993) and common field trial experiments containing range-wide provenance collections of *E. globulus* have shown that the timing to first flowering (Chambers *et al.* 1997) and flowering time period (Gore and Potts 1995) show considerable provenance variation. Dwarf ecotype trees undergo early vegetative phase change and precocious flowering relative to tall ecotype populations and the dwarf population from Wilsons Promontory in south-eastern Victoria is the most extreme of these dwarf populations (Dutkowski and Potts 1999; Jordan *et al.* 2000). For example, Wilsons Promontory *E. globulus* trees have been observed to produce adult foliage and flower buds within the first year of growth, whereas trees from later developing populations from northeast Tasmania populations may still be vegetatively and reproductively juvenile after four years (Jordan *et al.* 2000). Quantitative genetic analyses have shown that both the timing to vegetative phase change (Dutkowski and Potts 1999) and timing to first flowering (Chambers *et al.* 1997) are highly heritable in *E. globulus* and it is reported that these developmental processes are genetically independent in eucalypts (Jordan *et al.* 1999; Wiltshire *et al.* 1998). It is argued that the independence of these processes have enabled heterochronic change (change in the timing or rate of developmental processes; McKinney and McNamara 1991), an important mode of evolution, to occur within *Eucalyptus* species. This is evident by a number of independent speciation events having occurred within the genus that have involved a shift in flowering from the adult to juvenile vegetative phase (Potts and Wiltshire, 1997).

While a number of quantitative genetic studies have examined vegetative phase-change and flowering processes in eucalypts, little is known about the molecular genetic control of these processes. However, it is likely that both vegetative and reproductive phase-change transitions may involve similar, and equally complex, molecular pathways to those characterised in model annual plants. Most of what is known about the molecular genetics of flowering comes from the annual plant *Arabidopsis thaliana* (Henderson and

Dean 2004; Simpson *et al.* 1999). In *Arabidopsis*, the initiation of flowering is carried out through the syntheses of four genetic pathways (gibberellin, autonomous, vernalisation, and light-dependent pathways) which respond to various physical, chemical, and biological signals. Genetic signals from these pathways are part of an integrated pathway, which in turn transmits a signal to floral identity genes; thereby inducing floral morphogenesis (reviewed in Komeda 2004). Homologues of *Arabidopsis* genes involved in the flowering pathway have been identified in tree species, including *Eucalyptus* (Jaya *et al.* 2010b; Jones 2009; Kyozuka *et al.* 1997; Southerton *et al.* 1998a; Southerton *et al.* 1998b), and in many cases their function has been supported by transgenic complementation of *Arabidopsis* mutants. The presence of *Arabidopsis* homologues in a wide range of plants suggests that the general *Arabidopsis* flowering framework may be evolutionary conserved to some extent (Bäurle and Dean 2006; Hecht *et al.* 2005; Jones 2009). *Arabidopsis* has also been the model species of choice for studying vegetative phase change. Recent reports indicate that genome-encoded single-stranded RNA molecules of ~20 nucleotides in length known as micro-RNA (miRNA; Teune and Steger 2010) play a pivotal role in this transition (Chen *et al.* 2010; Wu *et al.* 2009b). Micro-RNAs mediate vegetative phase change through the negative regulation of target transcription factors (Wang *et al.* 2011). This occurs through complementary base pairing of miRNA to messenger RNA (mRNA) in the cell cytoplasm which leads to the degradation or translation inhibition of the mRNA (Teune and Steger 2010). In particular, action of *miRNA156* and *miRNA172*, and their interaction with *SQUAMOSA* Promoter-Binding Protein-Like (SPL) and *APETALA2*-related transcription factor targets, respectively, has been shown to regulate the juvenile-to-adult vegetative transition in *Arabidopsis* (Wu *et al.* 2009b). These processes appear to be conserved in grasses (rice and maize), and most recently, *miRNA156* has been shown to regulate vegetative phase change in *E. globulus* and other perennial trees (Wang *et al.* 2011).

While adaptation can be readily demonstrated in reciprocal transplant experiments (Lowry *et al.* 2009; Verhoeven *et al.* 2008), the relationship between adaptive traits and their underlying genetic architecture (e.g. the number, genomic location, effect and interactions of genes affecting these traits) remains poorly understood (Lowry *et al.* 2009; Remington and Purugganan 2003; Rogers and Bernatchez 2007; Verhoeven *et al.* 2008). Quantitative trait loci (QTL) mapping can play an important role in understanding the genetic architecture of population differentiation and trait variation

(Anderson *et al.* 2011; Stinchcombe and Hoekstra 2007; Verhoeven *et al.* 2008) and a number of recent studies have used this approach to detect QTL underlying adaptive traits, including; tolerance to salt spray and leaf sodium concentration between coastal and inland populations of monkeyflower (*Mimulus guttatus*; Lowry *et al.* 2009), flowering time variation between wild barley populations (*Hordeum spontaneum*; Verhoeven *et al.* 2008), cold-hardiness in Douglas-fir (Wheeler *et al.* 2005), frost tolerance in *E. nitens* (Byrne *et al.* 1997) and adaptive morphological traits between bladder campion ecotypes (*Silene vulgaris*; Bratteler *et al.* 2006).

While numerous studies have presented QTL for adaptive traits in forest trees, identifying the gene(s) underlying QTL variation remains difficult. Candidate loci that are potentially associated with trait differentiation may be identified by investigating whether known genes involved in the regulatory or biochemical pathways associated with trait expression co-locate with QTL (Pflieger *et al.* 2001; Thamarus *et al.* 2002). Recent developments in *Eucalyptus* genetics now make the prospect of identifying such candidate genes more feasible than ever. For example, the development of a Diversity Array Technology (DArT) genotyping array for *Eucalyptus* (Sansaloni *et al.* 2010; Steane *et al.* 2011) has facilitated the construction of high density genetic linkage maps which can be used for QTL detection. Additionally, most DArT markers contained on the eucalypt array have been sequenced enabling their position on the recently released assembled *E. grandis* genome (V1.0 released January 2011; www.phytozome.com) to be examined. Furthermore, the sequence information from the *E. grandis* genome can be transferred to related species (including *E. globulus*) due the high genome homology between *E. grandis* and other closely related *Symphyomyrtus* species (this thesis Chapter 1; Hudson *et al.* in press; Myburg *et al.* 2003). These recent developments will greatly assist in the identification of positional candidate genes for detected QTL.

In this study, QTL analysis was used to examine the genetic architecture underlying the phenotypic traits which differentiate dwarf and tall ecotypes of *E. globulus*. These ecotypes differ significantly in their timing to vegetative and reproductive phase-change and due to the large F_2 family of this study being planted at two field trials, it was possible to examine the stability of detected QTL across environments for these traits.

3.2 Materials and Methods

3.2.1 Genetic material and trial design

A large outcrossed F_2 family (termed Lighthouse F_2) was generated in order to study genetic variation between dwarf and tall ecotypes of *E. globulus.* The F_2 was created by crossing dwarf ecotype and tall ecotype grandparents which differed substantially in their timing to vegetative and reproductive phase change (Dutkowski and Potts 1999). Two dwarf ecotype grandparents (614LH and 615LH) from Wilsons Promontory (WP; southeast Victoria) were crossed with tall ecotype trees from King Island (Bass Strait Island) and Taranna (south-east Tasmania) provenances, respectively (Table 3.1). Two individuals (BA0010 and BA0012) from the resulting F_1 families were crossed to generate the F_2 family (Table 3.1).

Table 3.1 Pedigree information of the Lighthouse F_2 family, showing grandparent and F_1 generation family names and cross details; female (F), male (M) parent in cross.

Family	Ecotype	Population	Cross details	
			Tree ID	Progeny
Grandparents				
DFC219	Dwarf	Wilsons Promontory	614LH (F)	$F_1$960121 family
UTAS440	Tall	King Island	KI440 (M)	
DFC220	Dwarf	Wilsons Promontory	615LH (F)	$F_1$960124 family
UTAS423	Tall	Taranna	TA423 (M)	
F_1 *generation*				
$F_1$960121	-	-	BA0010 (F)	F_2 family
$F_1$960124	-	-	BA0012 (M)	

Seed of the F_2 generation were sown in April 2006 with seedlings raised in a glasshouse for 6 – 12 months before being planted in two field trials at Geeveston and Boyer in southern Tasmania. The Boyer field site is situated in the Derwent Valley approximately 22 km northwest of Hobart, whereas the Geeveston field site is some 50 km southwest of Hobart. Despite the trials being only roughly 50 km apart, they were quite different in terms of their climatic conditions (Table 3.2), in addition to altitude, slope, aspect and edaphic characteristics. The Geeveston trial was planted 6 months prior to Boyer. An irrigation system was installed at Boyer with watering applied as necessary. Trees were planted in a randomised block design and at 4 m x 4 m spacing (double normal plantation spacing in Tasmanian plantations) to avoid shading from neighbouring trees.

Table 3.2 Site and climate characteristics of Geeveston and Boyer trials.

Site and climate characteristic	Boyer	Geeveston
Site		
Coordinates	42º46´ 147º05´	43º13´ 146º53´
Planted	May – July 2007	November 2006
Altitude (metres above sea level)	10	350
Climate – 30 year (1981-2010) annual average[a]		
Annual rainfall (mm)	548	887
Mean annual maximum temperature (ºC)	18.0	16.8
Mean annual minimum temperature (ºC)	5.7	6.1

[a]From Australian Bureau of Meteorology weather stations closest to trial sites; Bushy Park (17 km from Boyer trial) and Geeveston (6 km from Geeveston trial).

A total of 563 trees across all families were scored for phenotypic traits across both sites (Table 3.3). This included 11-22 trees from each of the four grandparental open-pollinated families and the F_1 family $F_1$960121; only two trees were scored in the F_1 family $F_1$960124 as few trees of this family were planted (Table 3.3). Four hundred and sixty-seven F_2 individuals were used for QTL analyses. This comprised 158 and 309 F_2 individuals from Boyer and Geeveston trials, respectively. A greater total number of trees were scored at Geeveston (361) relative to Boyer (202). This was principally due to the Geeveston trial being planted earlier than the Boyer trial and suitable quantities of leaf material for DNA extractions becoming available on Geeveston trees earlier.

Table 3.3 Total number of trees per family scored and analysed for phenotypic traits at Boyer and Geeveston field sites.

Generation and family name	Boyer	Geeveston	Total
Dwarf ecotype grandparents			
DFC219[a] (Wilsons Promontory)	11	11	22
DFC220[a] (Wilsons Promontory)	9	12	21
Tall ecotype grandparents			
UTAS440[a] (King Island)	9	13	22
UTAS423[a] (Taranna)	4	7	11
F_1 generation			
$F_1$960121	10	8	18
$F_1$960124	1	1	2
F_2 family	158	309	467
Total	202	361	563

[a]family derived from open-pollinated seed collected from the grandparent tree.

3.2.2 Assessment of phenotypic traits

Two traits were scored prior to field planting on glasshouse seedlings at age 6 months; total number of fully expanded nodes (GH-EX) and node 0 to 7 length (GH-07). The last fully expanded node was marked with Color Tool florist spray-paint (Design Master, Colorado USA) for future scoring reference. Nodes were marked with flagging tape and paint in subsequent field trial assessments. Traits measured in field trials could be broadly classified into three groups; vegetative and reproductive phase change traits, flowering traits, and tree height and tree shape traits (Table 3.4).

3.2.2.1 Vegetative and reproductive phase change traits

Vegetative and reproductive phase change traits were scored at approximate six month intervals from planting until December 2010. At that time, Boyer and Geeveston trials were aged 3.5 and 4 years, respectively, and all F_2 trees had produced adult vegetation. In *E. globulus,* it is thought that vegetative phase change is initiated as a function of the number of nodes set rather than absolute age or size (Jordan *et al.* 1999). Therefore, node counts were made to the positions at which both vegetative and reproductive phase-change occurred. Juvenile, intermediate and adult leaf morphology classes were used to assess the node of vegetative phase-change (NJ; Table 3.4). Juvenile *E. globulus* leaves are sessile and the development of a small leaf petiole signalled the initiation of the juvenile-to-adult vegetative phase-change and the start of intermediate foliage in this study (Figure 3.1). Following this, plants proceed through an intermediate phase in

which foliage exhibits morphological characteristics of both juvenile and adult states. Leaves were assessed to be fully adult (NA; Table 3.4) when they had obtained large petioles and were highly pendulous in shape relative to intermediate foliage (Figure 3.1). Node counts were made on the main stem where possible, however, in highly branched trees where there was no clear apical dominance, or when foliage on the main stem had abscised, node of phase-change measurements were made on lateral branches.

For sexually reproductive trees, the node of first flower bud (NB; Table 3.4) was scored. Flower buds were mostly scored on lateral branches as they rarely developed on the main stem. When buds initiated on multiple branches in the canopy, multiple NB counts were made and the smallest value was taken. In these situations, node of first bud initiation was typically found to occur at the same node number regardless of canopy position. For all node counts made on lateral branches, nodes were firstly counted along the main stem from the cotyledons to the lateral branch position and then along the lateral branch. The height of phase-change transition (e.g. H-NJ, H-NA and H-NB; Table 3.4) was also recorded in this study; as height has been a more commonly scored character for vegetative phase-change in eucalypt studies (e.g. Freeman 2006; Jordan *et al.* 1999; Jordan *et al.* 2000). For nodes measured on the main stem, height was measured from the ground to this position. For nodes scored on lateral branches, height was scored as a cumulative measure of, (1) height from ground to lateral branch, and (2) along lateral branch from the main stem to the node scored. Mean juvenile internode length (Juv-Int-L; Table 3.4) and the number of intermediate foliage nodes (Int-N) were calculated from node and height to phase-change measurements (see Table 3.4).

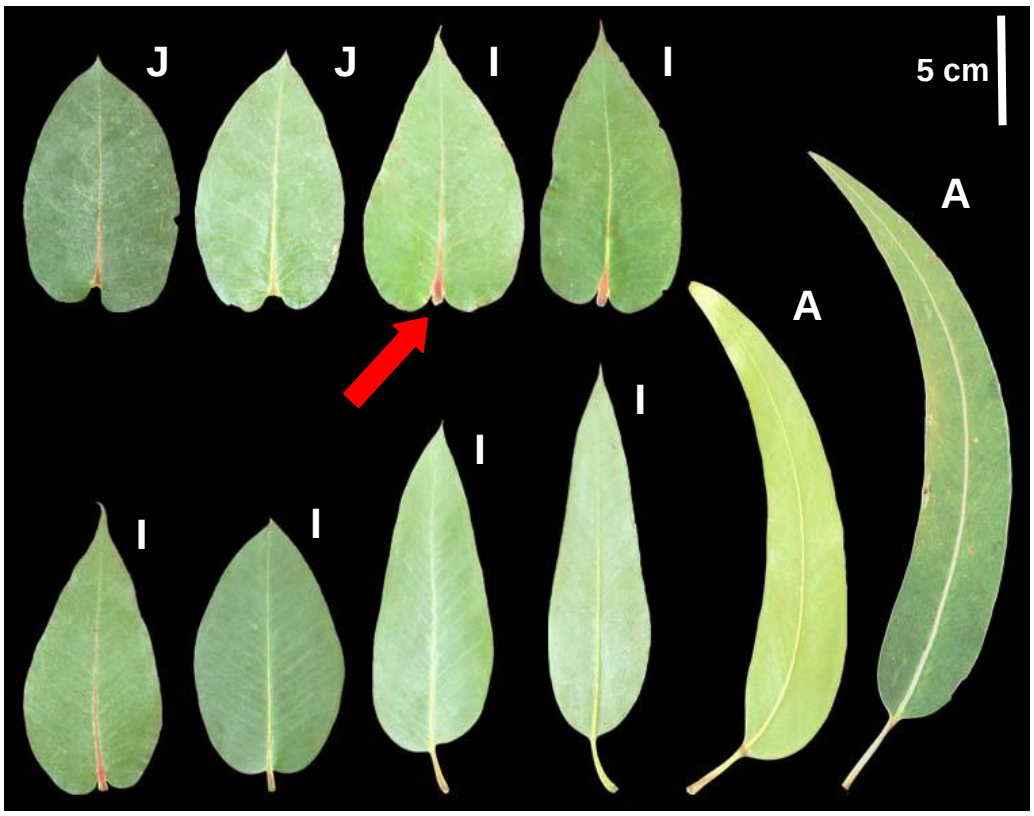

Figure 3.1 Foliage heteroblasty in *E. globulus*, showing the transition between juvenile (J; top left) to adult leaves (A; bottom right). Arrow indicates the emerging petiole on the first intermediate (I) leaf, indicating the initiation of vegetative phase-change.

Table 3.4 Description of traits measured in the glasshouse and field-trials.

Trait	Abbreviation	Description
Glasshouse measurements (6 months of age)		
Total number of fully expanded nodes	GH-EX	Number of fully expanded nodes at 6 months of age. Cotyledons scored as node 0
Node 0-7 length	GH-07	Length (cm) between nodes 0 to 7
Field trial measurements		
Vegetative and reproductive phase-change traits		
Node of last juvenile leaf	NJ	Node number of the last juvenile leaf; marking the vegetative phase-change transition
Height to node of last juvenile leaf	H-NJ	Height (cm) to NJ
Mean juvenile internode length	Juv-Int-L	Calculated as: H-NJ / NJ
Number of intermediate foliage nodes	Int-N	Calculated as: NA - NJ
Node of first adult leaf	NA	Node of the first fully adult leaf
Height to node of first adult leaf	H-NA	Height (cm) to NA
Node of first flower bud	NB	Node of first flower bud initiation
Height to node of first flower bud	H-NB	Height (cm) to NB
Flowering traits		
Median flowering time	MFT	Reported as the number of days from 1st April in the year of flowering
Flowered by 2010 / 11	F 10/11	A binary trait used to assess flowering precocity; flowered by the end of the 2010 flowering year = 1, not flowered = 0
Height and tree shape traits		
Tree height (June 2010 measurement)	H	Tree height as at June 2010
Height to widest point ratio (tree shape)	H/WP	Ratio of tree height / tree widest point (horizontal plane). Tree height and width measurements used for this ratio calculation were scored at both trials at age 23 months

3.2.2.2 Flowering traits

Across sites and years, flowering commenced in July (winter) and continued until January (summer) of the following year. For simplicity, flower time periods which spanned the calendar years 2009/10 and 2010/11 are referred to as 2009 and 2010 flowering years, respectively. Flowering trees were scored at weekly or fortnightly intervals throughout each flowering time period at both sites. The percentage of flower buds in flower, having already flowered, and remaining to flower were recorded. This data was used to estimate median flowering time (MFT), which is reported as the number of days from April 1 in each year and was estimated from scale and shape parameters derived by fitting a two-parameter Weibull function to the measure of 'percent bud crop remaining to flower' using PROC NLIN in SAS (Version 9.2; SAS Institute, Cary, USA). Buds were scored as 'being in flower' from the time of flower bud operculum-lift (*i.e.* anthesis) until flower stamens were withered and had begun to abscise. Flowering precocity at the end of the 2010 (F 10/11) was assessed by transforming NB data. Plants were qualitatively classified as precocious (1 = buds produced) or non-precocious (0 = no buds produced) using NB data regardless of whether the bud developed into a mature flower. As 78% of Boyer F_2 individuals had flowered by the end of the 2010 flowering year, a second individual-site QTL analysis was conducted for this site using 2009 flowering precocity data; 22% of F_2 trees had flowered at this time.

3.2.2.3 Tree height and shape characters

The final tree height (H) measurement taken for all trees in June 2010 was used for QTL analysis. Tree shape measurements were scored at both field sites at 23 months after planting. At this time, tree height (H) and widest point of the canopy (WP; measured on the horizontal plane) were measured. A height to width ratio (H/WP) was calculated to reflect tree shape; with tall-thin trees having a larger H/WP ratio relative to short-wide trees.

3.2.3 Analysis

QTL analyses were performed on F_2 individuals from Boyer and Geeveston trials in individual-site analyses and also using data combined across-sites, providing an opportunity to examine the stability of QTL across sites. For each individual-site analysis and for combined-site analyses in which data had been site-adjusted (mean of zero, standard deviations different, performed in SAS 9.2 using PROC STANDARD),

trait distributions were examined in order to test the assumption of normality for interval and restricted multiple-QTL mapping (rMQM) mapping. Most traits in both individual-site and combine-site analyses departed from normality (Kolomogorov-Smirnov α 0.05), however, for the majority the data were evenly distributed around the mean and therefore no transformations were undertaken. Pearson's correlation coefficients were calculated between traits using PROC CORR of SAS. Contrast tests were conducted to test for significant differences in trait values between dwarf and tall ecotype grandparental families (DFC219 and DFC220 *vs.* UTAS440 and UTAS423). In each test a mixed model was fitted with site, family, and their interaction as fixed effects and replicate within site as a random effect using PROC MIXED in SAS.

QTL analyses were primarily conducted using MapQTL 6.0 (Van Ooijen 2009). Genotypic data and map information from the *E. globulus* Lighthouse F_2 sex-averaged consensus linkage map described in Chapter 1 of this thesis was utilised for MapQTL analysis. Briefly, individual parental maps were built using 503 F_2 individuals before constructing an integrated consensus map in JoinMap 4.0 (Van Ooijen 2006). This 1060-marker map contained 50 microsatellite (simple sequence repeats; SSRs) and 1010 DArT molecular markers and totalled 1151 cM over 11 linkage groups. The numbering of linkage groups followed the established order of Brondani *et al.* (2006) and also corresponded to the numbering of the *E. grandis* genome sequence chromosome assemblies (www.phytozome.net/). Due to the high computational demands of using dominant marker types such as DArT in multiple-QTL mapping (MQM) (Van Ooijen 2009), a reduced-marker map was required in order to perform restricted-MQM (rMQM) QTL analyses. A reduced-marker linkage map containing 391 markers was obtained by simply removing most 3:1 segregating DArT markers, and by retaining an even distribution of DArT markers segregating in a 1:1 ratio from both parents at approximate 2-5 cM intervals only from the 1060-marker map; all SSRs were retained. Marker linkage-group positions from the 1060-marker map were used for 391-marker map QTL analyses (*i.e.* recombination values were-not recalculated for the 391-marker map).

In interval mapping (IM), the likelihood for the presence of a segregating QTL is calculated for each position in the genome (*i.e.* on a linkage map) and during this procedure the genotypic effects of QTL genotypes (*i.e.* qq, Qq and QQ) are calculated along with a residual variance (Van Ooijen 1992). This likelihood is compared to the likelihood under the null-hypothesis (there is no segregating QTL) and a likelihood ratio

statistic called the LOD (or LOD score) is calculated; which is the 10-base logarithm of the quotient of the two respective likelihoods (Van Ooijen 2009). When calculating genotypic effects for a given genome position in a single QTL model (e.g. IM) the residual variance component is effected by other QTL affecting the same trait (Van Ooijen 1992). Therefore, by selecting markers linked to QTL as cofactors in MQM mapping (a similar method to composite interval mapping; Van Ooijen, 2009) the residual variance is removed; thereby increasing the power and precision for subsequent QTL searches. Therefore, following the detection of putative QTL in IM, markers linked to QTL which exceeded the suggestive threshold (α 0.05) were chosen as cofactors in rMQM mapping. In rMQM mapping, all cofactor markers, except those on the linkage group in which the QTL is being fitted, are used in the modelling of genotypic effects. rMQM analyses were performed in an iterative approach until no further QTLs were detected, selected cofactor markers were the closest marker to each QTL, and QTL positions were stable.

All rMQM analyses were conducted on the 391-marker map for both individual-site and combined-site analyses. In combined-site analyses 'site' was fitted as an experimental design cofactor in MapQTL 6.0. Permutation tests were run in MapQTL 6.0 to determine LOD significance thresholds at genome-wide and chromosome-wide levels for all traits (1000 permutations; Churchill and Doerge 1994). Putative QTL were declared at three significance thresholds, significant QTL (genome-wide type I error $\alpha \leq 0.05$), suggestive QTL (chromosome-wide type I error rate $\alpha \leq 0.05$) and suggestive QTL at a higher chromosome-wide type I error rate ($0.05 < \alpha < 0.10$; termed 'possible' QTL). To determine 95% QTL confidence intervals, the two positions to the left and right of the QTL LOD peak with a value of two less than the maximum LOD were taken (Van Ooijen 2006). Regression approximation to the maximum likelihood algorithm implemented in MapQTL 6.0 was used in interval and rMQM mapping. Single marker Kruskal-Wallis tests were also performed on combined-site data for QTL detection. These tests were used to determine from which parent(s) QTL effects segregated. QTL positions were mapped using MapChart 2.2 (Voorrips 2002).

To detect QTL by environment (site) interaction, a mixed model was fitted using PROC MIXED in SAS for markers flanking each QTL detected in combined-site rMQM analyses. Site, marker, and their interaction were fitted as fixed effects and replicate within site as a random effect in the model. A Tukey adjustment was applied when testing for differences between site by QTL genotype least squares means.

A second QTL analysis method based on Statistical Machine Learning (SML; Bedo *et al.* 2008) was also performed. The SML method is conceptually quite different to QTL interval-mapping in that it estimates the generalisation performance of a QTL model by splitting trait data into independent training and testing subsets that are used for model induction and evaluation, respectively; a full description of the program can be found in Bedo *et al.* (2008). SML QTL analyses involve single marker tests which do not consider the map position of the markers tested. Therefore, un-mapped markers which had been discarded during the construction of the *E. globulus* Lighthouse F_2 consensus map could be included in SML analyses. From the original DArT marker dataset generated in this F_2 family, DArT markers with a call rate > 80 and *P* value > 65 (see Jaccoud *et al.* 2001) were included in SML analyses; this totalled 2047 DArT markers. For DArT markers found to have a significant SML-QTL association that were not mapped on the *E. globulus* Lighthouse F_2 consensus linkage map, the *Eucalyptus* multi-species composite map described in Chapter 2 of this thesis was used to estimate marker position in order to facilitate comparison of SML and MapQTL results. QTL detected with MapQTL were considered validated by SML if the significant ($\alpha \leq 0.05$) QTL-associated DArT marker resided within the MapQTL 2-LOD confidence interval.

Following the identification of QTL, BLAST servers available at BOGAS (http://bioinformatics.psb.ugent.be/webtools/bogas/) and Phytozome (http://www.phytozome.net/eucalyptus.php) were used to search the *E. grandis* genome for eucalypt homologues of candidate genes (or proteins) known to be involved in the genetic control of vegetative phase-change for which highly significant QTL were detected. These were *JAT* (Fernández-Ocaña *et al.* 2010) and *DAL1* (Carlsbecker *et al.* 2004) genes, and *Arabidopsis* micro-RNA *miRNA156a*, *SPL* family (1-12, 13A, 13B and 14-16) and *AP2*-related (*TOE1*, *TOE2*, *TOE3*, *SMZ* and *SNZ*) gene and protein sequences (obtained from http://www.arabidopsis.org/; Chen *et al.* 2010; Wang *et al.* 2011; Wu *et al.* 2009b; Wu and Poethig 2006).

3.3 Results

Comparison of dwarf and tall ecotype grandparental family trait values using combined-site data showed that grandparental ecotypes differed significantly (all $P < 0.05$) for all measured traits; apart from one glasshouse measurement (GH-EX, *P* value 0.22, F 1.5; Table 3.5). Wilsons Promontory dwarf ecotype grandparental families (DFC219 and DFC220) could be characterised as having slower growth, smaller internode elongation and less apical dominance (H/WP) in comparison to tall ecotype grandparents (UTAS440 and UTAS423 families). Although their growth rate was much slower, dwarf ecotype trees underwent a faster transition to both vegetative and reproductive maturity in comparison to tall ecotypes. For example, first flower bud development (NB) occurred at 36.9-39.5 nodes in dwarf ecotype grandparent families and almost all (38 out of 43) dwarf ecotype grandparent trees had flowered by the end of the 2010 flowering year. In contrast, only 2 out of 33 tall ecotype grandparent trees had flowered at this time, with first bud development occurring at nodes 65 and 69, respectively (Table 3.5). At December 2010 (final phenotypic scoring), dwarf ecotype grandparent trees typically had a full canopy of adult vegetation and had flowered, whereas most tall ecotype grandparent trees still contained a large proportion of juvenile foliage and only two trees had reached reproductive maturity. Mean trait values for grandparental, F_1 and F_2 families can be seen in Table 3.5; for simplicity only site-adjusted combined-site data is presented. Individuals of both Wilsons Promontory dwarf ecotype families (DFC219 and DFC220) were phenotypically very similar. However, some traits were detected to differ significantly between tall ecotype families (UTAS423 and UTAS440). For example, tall ecotype families differed in their node to vegetative phase-change transition (NJ, $P < 0.0001$), number of intermediate foliage nodes (Int-N, $P < 0.0001$) and node to first adult leaf (NA, $P < 0.0001$; Table 3.5).

Table 3.5 Combined-site (site-adjusted) family mean trait values and results of contrast tests between grandparental (GP) dwarf *vs.* tall ecotype families; KI=King Island, TA=Taranna. For GP families, superscript letters indicate significant family differences ($\alpha \leq 0.05$); n/a indicates test not applicable due to small sample size. The number of individuals per family and the cross details of the F_1 families are given under family names.

Trait and unit of measurement[a]		Family means							
		Dwarf GPs		Tall GPs		GP ecotype contrast *P* value[b]	F_1 Families		F_2 (467)
		DFC219 (22)	DFC220 (21)	UTAS 440(KI) (22)	UTAS 423(TA) (11)		$F_1$960121 (DFC219xKI) (18)	$F_1$960124 (DFC220xTA) (2)	
GH-07[c]	cm	110.9[a]	122.0[a]	176.1[b]	183.5[b]	0.0001	154.7	203.0	157.2
GH-EX[c]	N	9.9[a]	9.9[a]	10.6[a]	8.4[b]	0.23	10.7	9.0	8.5
H	cm	241.4[a]	204.0[a]	505.7[b]	438.3[b]	0.0001	385.7	436.5	390.1
NJ	N	21.5[a]	19.0[a]	56.5[b]	29.4[b]	0.0001	35.1	26.4	26.8
J-Int-L	cm	2.0[a]	2.0[a]	3.9[b]	3.5[b]	0.0001	2.8	3.4	2.9
Int-N	N	11.9[a]	13.2[ac]	16.4[bc]	25.3[b]	0.0001	14.2	17.0	15.5
NA	N	32.9[a]	32.4[a]	72.3[b]	55.8[c]	0.0001	49.4	43.5	42.4
NB	N	39.5[a]	36.9[a]	69.0[n/a]	65.0[n/a]	0.0001	57.0	40.0	48.2 [n/a]
MFT[d]	days (*n*)	248[a] (13)	239[a] (14)	251[n/a] (1)	214[n/a] (1)	n/a	236.0 (8)	184 (1)	238 (186)
F 10/11	%	86.4[n/a]	90.5[n/a]	4.5[n/a]	9.1[n/a]	0.0001	61.1	50.0	52.5
H/WP[e]	ratio	1.6[a]	1.7[a]	2.4[b]	2.2[b]	0.0001	2.2	2.5	2.0

[a]unit of measurement, N=node. [b]Contrast tests between dwarf ecotype families DFC219 and DFC220 *vs.* tall ecotype families KI440 and TA423. [c]measured in glasshouse at seedling age 6 months, no site-adjustment applied. [d]Mean flowering time (MFT) combined-site data calculated using Boyer MFT 10/11 and Geeveston MFT 09/10 data. MFT measurement unit is number of days from 1st April. The number (*n*) of trees which flowered in these years across sites is shown in parentheses. [e]Tree shape scored at 23 months trial age.

In the F_2 family, height to phase-change was highly correlated with node of phase-change (Table 3.6). For example, the last juvenile leaf node (NJ), marking the point of juvenile-to-adult vegetative phase-change, was significantly correlated (0.82; $P < 0.0001$) with height to NJ (H-NJ), and similarly, height to the first adult leaf node (H-NA) and height to first flower bud (H-NB) were also highly correlated (0.73 and 0.67, both $P < 0.0001$; Table 3.6) with their respective node counts. In preliminary interval mapping analyses, the same putative QTL were detected for 'height to phase change' traits (H-NJ, H-NA and H-NB; data not shown) as detected with their corresponding 'node of phase-change' traits (NJ, NA and NB). This suggested that both 'node' and 'height' measures captured much of the same phase-change information, however, LOD scores, percentage of variance explained (PVE) and the number of significant QTL detected were greater for 'node of phase-change' traits. Therefore, results have only been presented for node of phase-change traits, as they provided more information in comparison to height to phase-change traits (*i.e.* H-NJ, H-NA and H-NB).

Table 3.6 Pair-wise correlation coefficients for vegetative and reproductive phase-change, tree shape and median flowering time traits in the F_2 family; calculated on combined site-adjusted data. No site adjustment was applied to glasshouse traits (GH-07 and GH-EX). Descriptions of trait abbreviations are provided in Table 3.4.

	GH-07	GH-EX	H	NJ	H-NJ	J-Int-L	Int-N	NA	H-NA	NB	H-NB	H/WP
GH-07	-											
GH-EX	0.48^{****}	-										
H	0.12^{*}	0.05^{ns}	-									
NJ	-0.08^{ns}	0.27^{****}	0.09^{*}	-								
H-NJ	0.07^{ns}	0.24^{****}	0.32^{****}	0.82^{****}	-							
Juv-Int-L	0.25^{****}	0.17^{***}	0.43^{****}	0.35^{****}	0.79^{****}	-						
Int-N	-0.02^{ns}	-0.06^{ns}	-0.09^{ns}	-0.15^{**}	-0.21^{****}	-0.21^{****}	-					
NA	-0.10^{*}	0.20^{****}	0.02^{ns}	0.82^{****}	0.62^{****}	0.19^{****}	0.43^{****}	-				
H-HA	0.06^{ns}	0.14^{**}	0.41^{****}	0.65^{****}	0.77^{****}	0.61^{****}	0.24^{****}	0.73^{****}	-			
NB	-0.17^{**}	0.15^{*}	0.12^{ns}	0.64^{****}	0.42^{****}	-0.03^{ns}	0.22^{***}	0.73^{****}	0.47^{****}	-		
H-NB	0.07^{ns}	0.09^{ns}	0.44^{****}	0.32^{****}	0.42^{****}	0.31^{****}	0.09^{ns}	0.36^{****}	0.55^{****}	0.67^{****}	-	
H/WP	0.15^{**}	0.05^{ns}	0.28^{****}	0.10^{*}	0.19^{****}	0.18^{****}	0.05^{ns}	0.13^{**}	0.23^{****}	0.13^{*}	0.20^{**}	-
MFT	0.07^{ns}	-0.90^{ns}	-0.12^{ns}	-0.07^{ns}	-0.07^{ns}	-0.02^{ns}	-0.04^{ns}	-0.07^{ns}	-0.05^{ns}	0.02^{ns}	0.06^{ns}	-0.07^{ns}

^{ns} indicates a non-significant correlation ($\alpha > 0.05$), $^{*}P \leq 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ and $^{****}P < 0.0001$.

In MapQTL analyses, 22 significant (genome-wide $\alpha < 0.05$) putative QTL and a further 11 QTL at the suggestive LOD threshold (chromosome-wide $\alpha < 0.05$) were detected in combined-site rMQM analyses for the nine field-measured characters analysed (Table 3.7). QTL were detected on all 11 linkage groups and the number of QTL detected per trait ranged from 2 to 5 (Table 3.7). A number of QTL were found to co-locate to the same position within linkage groups (*i.e.* overlapping 2-LOD support intervals); nonetheless 20 out of 33 QTL mapped to unique positions (Figure 3.2). Most QTL were of relatively small effect (less than 5% of phenotypic variation explained; PVE), however, a major effect QTL detected on linkage group three (LG3) explained 62.8% of the phenotypic variation for node of last juvenile leaf (NJ; Table 3.7). Although most QTL were also detected in at least one individual-site analyses, the proportion detected for each trait class varied. For example, 17 out of 18 vegetative and reproductive phase-change trait QTL, and 7 out of 10 flowering QTL, were detected in individual-site analysis. In contrast, individual-site analyses detected only one out of five tree height and tree shape QTL detected in combined-site analyses. Compared to other traits classes, tree height and tree shape QTL were of lower significance and only one QTL exceeded the genome-wide significance threshold (Table 3.7). Six QTL were detected in individual-site analyses at both trials (chromosome-wide $\alpha < 0.05$). Three of these, being detected for NJ, NA and NB (all pair-wise highly correlated traits; see Table 3.6) co-located on LG3 (Table 3.7 and Figure 3.2). A larger number of individual-site QTL were detected in the larger Geeveston trial (F_2 n = 309, 18 QTL detected) in comparison to Boyer (F_2 n = 158, 13 QTL detected). However, when excluding the flowering precocity trait (F 10/11; all five QTL detected in individual-site analyses at Geeveston only), the number of individual-site QTL detected in both trials was the same.

Of the 33 QTL detected using MapQTL, 24 of these same QTL were detected using SML (Table 3.7). This included 20 out of 22, and 4 out of 11, of the QTL detected at significant (genome-wide $\alpha < 0.05$) and suggestive (chromosome-wide $\alpha < 0.05$) significance levels using MapQTL, respectively. For 17 of the 24 QTL detected with both methods, the DArT marker found to have SML-QTL association resided within the 1-LOD confidence interval of the MapQTL-detected QTL (data not shown); with the remainder being within the 2-LOD confidence interval. Unless specified, the following results section refers to MapQTL results on data combined across sites.

Table 3.7 QTL detected using rMQM mapping on data combined across sites; position, LOD, parent from which the QTL effect segregated (Seg.; M = male, F = female) and percent variation explained (PVE) is shown. Combined-site QTL detected in SML QTL analyses are indicated by Y=Yes, or, N=No. Individual site analysis indicates whether each combined-site QTL was detected in individual-site rMQM mapping analyses, respectively. *P* values are shown for tests of QTL x site interaction, trait description are given in Table 3.4.

Trait	QTL detected in combined site analysis							Ind. site analysis[c]	QTL x site interaction *P* value
	LG	cM	Adj. marker[b]	LOD	PVE	Seg.	SML		
Vegetative and reproductive phase-change traits									
NJ	2	32.4	ePt-641876	4.43***	1.6	F	Y	B**	0.72
	2	48.7	ePt-568767	4.72***	1.7	F	Y	B***	0.90
	3	62.1	ePt-639243	103.0***	62.8	Both	Y	G*** B***	0.06
	5	48.6	ePt-641489	4.43***	1.6	M	Y	B***	0.93
	5	65.1	ePt-571521	4.01**	1.4	Both	Y	G* B***	0.34
Juv-Int-L	3	66.8	ePt-640855	15.59***	10.4	Both	Y	G***	< 0.0001
	4	10.9	ePt-600106	3.82**	2.5	M	N		0.66
	4	59.6	ePt-570676	5.8***	3.8	F	N	G***	0.61
Int-N	4	27.9	Es54	5.43***	3.8	Both	Y	G**	0.71
	6	34.8	Embra627	3.48**	2.5	F	N	B*	0.06
	11	62.1	ePt-575083	9.06***	6.5	Both	Y	G***	0.79
NA	2	47.8	ePt-641539	5.33***	2.0	F	N	G** B**	0.44
	3	62.8	ePt-639243	80.39***	44.9	Both	Y	G*** B***	0.45
	4	34.0	ePt-565545	7.06***	2.6	Both	Y	G***	0.47
	5	45.6	ePt-569568	4.13**	1.5	Both	Y	B**	0.84
	11	60.1	ePt-575083	9.81***	3.7	Both	Y	G***	0.27

NB	1	3.1	Embra11	3.29**	3.7	M	N	B*	0.35
	3	61.0	ePt-639927	15.87***	20.2	Both	Y	G*** B***	0.60
Flowering traits									
F 10/11	4	66.9	ePt-568492	9.38***	5.2	F	Y	G***	0.0044
	7	65.3	ePt-504063	4.3**	2.4	F	N	G***	0.57
	8	44.6	ePt-638446	5.43***	3.1	M	Y	G***	0.85
	8	78.7	ePt-640315	8.22***	4.6	Both	Y	G***	0.09
	10	70.9	ePt-572657	4.73***	2.6	M	Y	G**	0.47
MFT	3	73.0	ePt-570139	4.66***	2.1	F	Y		0.24
	4	72.4	Embra36	5.19***	2.4	Both	Y	B**	0.013
	5	0.9	Embra618	3.36**	1.5	M	N		0.32
	6	94.7	ePt-504481	5.23***	2.4	F	Y		0.0392
	8	129.5	Es76	7.98***	3.8	Both	Y	G***B***	0.0023
Tree height and shape traits									
H	3	42.4	ePt-571733	3.15**	2.7	F	Y	G**	0.54
	11	15.2	ePt-570063	3.56**	3.1	F	N		0.91
	11	39.5	Eg99	3.52**	3.0	F	N		0.08
H/WP	4	54.2	ePt-564417	3.4**	3.0	F	Y		0.24
	9	30.2	ePt-505052	4.44***	3.9	M	Y		0.16

[a]QTL LOD peak position on 391-marker map. [b]Adjacent marker (Adj. marker; 391-marker map) to QTL LOD peak. [c]Indicates if combined-site QTL was detected in individual-site analyses at Boyer (B) and Geeveston (G). LOD significance is indicated for both combined and individual site QTLs; ***genome-wide $\alpha \leq 0.05$; **chromosome-wide $\alpha \leq 0.05$; *chromosome-wide $0.05 < \alpha \leq 0.10$.

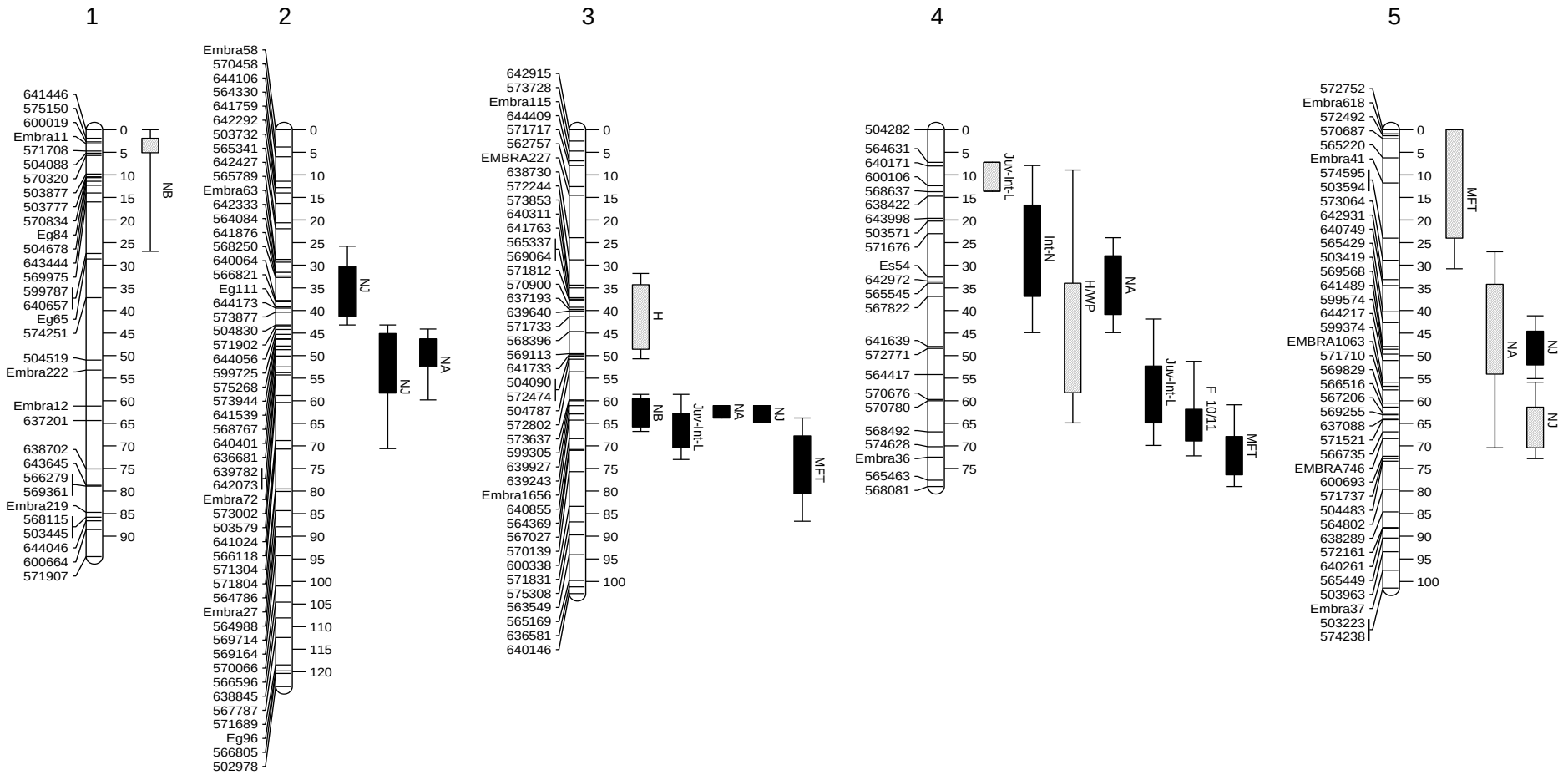

Figure 3.2 The location of putative QTL exceeding genome-wide significance ($\alpha \leq 0.05$; solid bars) and chromosome-wide significance ($\alpha \leq 0.05$; dashed bars) in combined-site rMQM MapQTL analysis. Loci of the *E. globulus* Lighthouse F_2 391-marker map are shown (see materials and methods). Bars and whisker symbols represent one and two-LOD support intervals, respectively. Descriptions of trait abbreviations are given in Table 3.4.

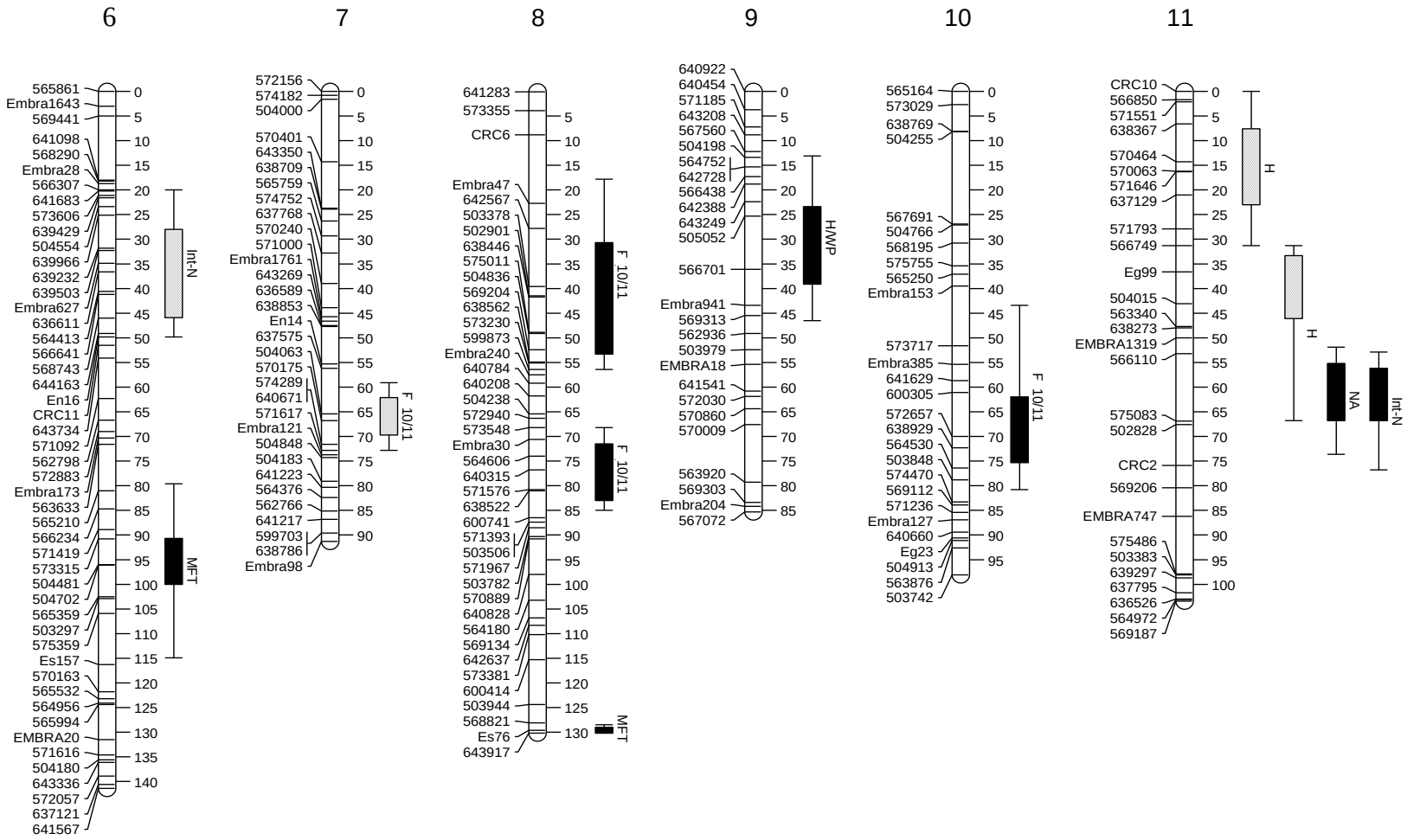

Figure 3.2 continued.

A significant QTL by site interaction was detected for the juvenile internode length QTL detected on LG3 (Juv-Int-L, $P < 0.0001$; Table 3.7). In the F_2 family, juvenile internode lengths were found to differ significantly between trials ($P < 0.0001$), with greater internode elongation occurring in plants at Geeveston (mean 3.2 cm) in comparison to Boyer (mean 2.3 cm). Despite this internode length variation in the juvenile vegetative phase, vegetative phase-change was initiated at the same node in the F_2 family in both trials; NJ mean in both trials 27.0, $P = 0.69$. In contrast, significant differences between trials were detected in both node to first adult (NA mean values; Boyer 45.5, Geeveston 40.0, $P < 0.0001$) and node to first flower bud traits (NB mean values; Boyer 52.3, Geeveston 43.0, $P < 0.0001$).

Three highly significant QTL detected for NJ (LOD 103), NA (LOD 80) and NB (LOD 16) co-located at ~62 cM on LG3 (Table 3.7 and Figure 3.2). These QTL were highly significant in both individual-site analyses and were also detected using SML. A microsatellite, Embra1656, was mapped within the confidence interval of these QTL (Figure 3.2 and Figure 3.3). Mean NJ, NA and NB values calculated for the four genotype classes of this microsatellite indicated that this QTL strongly influenced the node of transition to both vegetative and reproductive phase-change. For example, F_2 trees which inherited both dwarf ecotype grandparent alleles (genotype FG; Table 3.8) commenced vegetative phase-change at node 19, developed adult foliage at node 35 and produced flower buds at node 42 (mean values; Table 3.8). In contrast, trees which inherited both tall ecotype Embra1656 alleles (genotype EE; Table 3.8) underwent phase change at a significantly later number of nodes (NJ, NA and NB all $P < 0.0001$), and trees containing 1-dwarf and 1-tall Embra1656 allele had values which were intermediate between dwarf and tall grandparental classes (Table 3.8). Contrast tests between pooled dwarf and tall ecotype homozygous (e.g. FG and EE) and pooled heterozygous genotype class values (e.g. FE and GE) showed that the alleles of this microsatellite essentially had an additive effect on the traits, since there was no dominance (contrast test *P* values; NJ 0.72, NA 0.10 and NB 0.94).

Table 3.8 Mean trait values for microsatellite Embra1656 genotype classes in the F_2 family. Means are shown for last juvenile node (NJ), node to first adult (NA) and node to first bud (NB). 439 F_2 Embra1656 genotypes were obtained; there were 28 (6%) missing genotype scores. Grandparental (GP) ecotype alleles (*i.e.* dwarf or tall) are shown for each genotypic class (FG, FE, EG and EE). Superscript letters following the mean trait value indicates significance (adjusted Tukey α 0.05) between genotype classes within traits. For node to bud, the percentage of trees which had produced flower buds by the end of the 2010 flowering year is shown for each genotype class in parentheses.

	Embra1656 F_2 family genotypic class			
Alleles	FG	FE	GE	EE
n	126	129	89	95
GP alleles and ecotype[a]	DFC220 DFC219 2 x dwarf	DFC220 KI440 1 dwarf, 1 tall	DFC219 TA423 1 dwarf, 1 tall	KI440 TA423 2 x tall
Last juvenile node				
n (total 428)	123	126	88	91
Average	18.8[a]	28.1[b]	27.0[b]	35.8[c]
Node of adult				
n (total 428)	120	127	88	93
Average	34.8[a]	43.4[b]	41.3[b]	51.7[c]
Node to bud				
n (total 228)	64 (50.8%)	73 (56.6%)	45 (50.6%)	46 (48.4%)
Average	42.4[a]	48.5[b]	49.1[b]	55.7[c]

[a]Indicates which grandparent of the F_2 cross the allele came from, DFC219/20 are dwarf ecotype grandparents, TA423 and KI440 are tall ecotype grandparents.

Markers mapped within the 2-LOD interval of the major QTL detected on LG3 for NJ, NA and NB traits can be seen in Figure 3.3. As determined by BLASTn, markers within the NJ 2-LOD QTL confidence interval (61.1 - 64.8 cM) spanned a ~2 Mbp genomic region on LG3 of the *E. grandis* genome sequence (scaffold 3: 48,270,197..50,200,599; http://www.phytozome.net/eucalyptus.php). *Eucalyptus grandis* genome searches were performed using BLASTn (nucleotide sequences) or TBLASTx (protein sequences) for candidate genes known to be involved in vegetative phase change processes. A strong candidate gene for the control of vegetative phase-change, an *Arabidopsis miRNA156a* homolog (named *EgMIR156C*; Poethig *pers comm.*), was detected approximately ~ 1 cM from the QTL peak for NJ (scaffold 3: 50,822,715..50,822,629); no other candidate genes were found in this region. The estimated position of *EgMIR156C* relative to QTL detected for NJ, NA and NB on LG3 is shown in Figure 3.3.

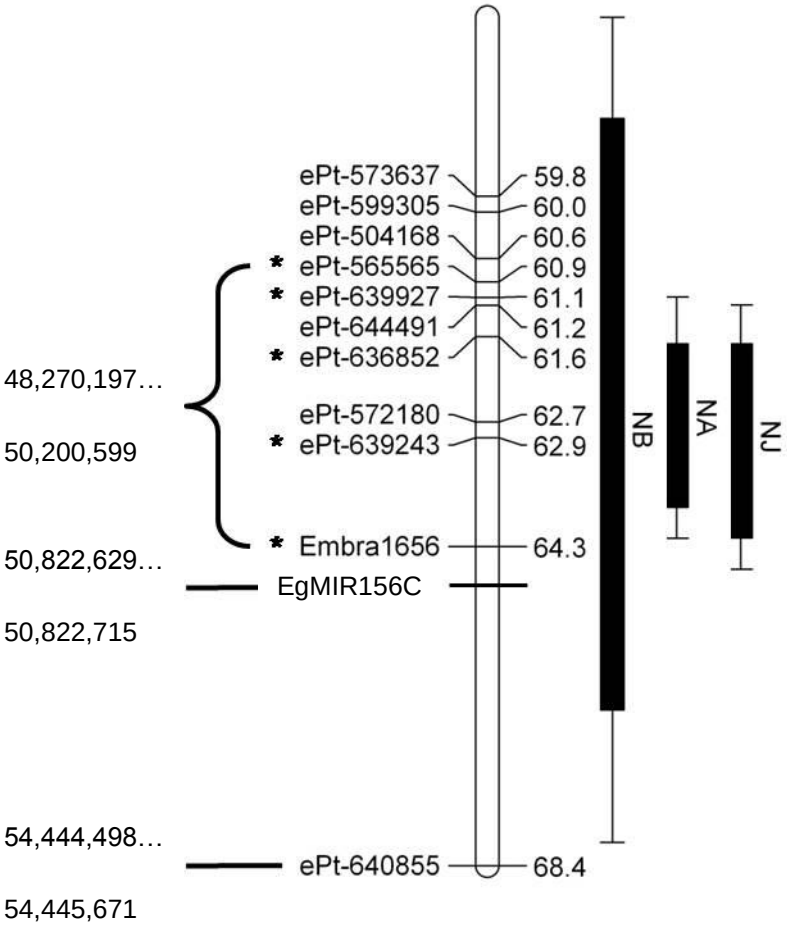

Figure 3.3 *E. globulus* Lighthouse F_2 consensus linkage group three region (59.8-68.4 cM) showing the positions of QTL detected for NJ, NA and NB in combined-site rMQM analyses; bar and whisker plots indicate 1-LOD and 2-LOD QTL confidence intervals. Markers mapped to the 1060-marker linkage map are shown. In the 391-marker map used for rMQM QTL analyses, marker ePt-639234 was the closet marker to the NJ LOD peak. The *E. grandis* genome positions (base-pairs) of markers are shown to the left of marker names. Five markers (indicated with asterisk) were located within the *E. grandis* genome region 48,270,197..50,200,599; marker ePt-572180 returned a BLAST hit to scaffold 2 and no sequence was available for marker ePt-644491. The approximate position of micro-RNA EgMIR156C is shown; estimated using the *E. grandis* genome position of EgMIR156C relative to other markers.

In addition to the major QTL detected on LG3 for NA, a further four QTL were detected for this trait. Three (on LG's 2, 3 and 5) and two (on LG's 4 and 11) of these QTL co-located with QTL detected for NJ and Int-N (number of intermediate nodes), respectively. Thus, the QTL detected for NA was a combination of genes (QTL) which influenced both the juvenile and intermediate developmental phases. Conversely, no QTL detected for NJ and Int-N co-located, suggesting that the genetic basis which

influences the duration of both juvenile and transitional developmental phases appears to be independent.

Although F_2 individuals which were homozygous for dwarf ecotype Embra1656 alleles produced flower buds at fewer nodes in comparison to individuals of other Embra1656 genotype classes, the percentage of F_2 trees in each Embra1656 class which had flowered by the end of the 2010 flowering year were similar (all approximately 50%; Table 3.8). Therefore, although the QTL detected on LG3 appears to strongly influence the node number at which vegetative and reproductive phase-change takes place, other genetic factors appear to influence if flowering occurs.

Two hundred and forty-two F_2 individuals had flowered by the end of the 2010 flowering year across both sites; 52% of all F_2's. Within trials, a larger percentage of F_2 individuals had flowered at Boyer (78%) relative to the Geeveston trial (33%). Five QTL were detected for flowering precocity (F 10/11; a binary scored trait) and all were found to be independent to QTL detected for NJ, NA and NB traits (Table 3.7 and Figure 3.2). A significant QTL by environment interaction was detected for the F 10/11 QTL detected on LG4 (marker ePt-568492, $P = 0.0044$; Table 3.7). Least squares mean estimates for the frequency of individuals having flowered in the two genotype classes of this QTL flanking marker were 0.88 and 0.90 at Boyer (non-significant genotype class difference, $P = 0.98$), and 0.27 and 0.54 at Geeveston ($P < 0.0001$). In individual-site analyses, this QTL, in addition to all other QTL detected for F 10/11, were detected at the Geeveston site only (Table 3.7). Furthermore, no QTL were detected in an individual-site analysis conducted for the Boyer trial using 2009 flowering year precocity data; at that time 22% of Boyer F_2 individuals had flowered (data not shown).

In both years in which flowering occurred at both sites (2009 and 2010), there was minimal overlap in site flowering time period (*i.e.* start and finish date of flowering). In both years, flowering commenced earlier and occurred over a greater time period at Boyer relative to the Geeveston trial; which commenced flowering later in the year and had a much shorter flowering time period (data not shown for all years). Although being different years, this trend can be seen when comparing Figure 3.4 (flowering at Boyer in 2010) and Figure 3.5 (flowering at Geeveston in 2009). Although only two tall ecotype grandparental trees flowered throughout the study period (both in 2010 at Boyer), and only a small number of dwarf-grandparental and F_1 family individuals flowered in this same year, the two tall grandparental trees flowered later in the flowering time period

compared to dwarf grandparental family trees and F_1 family trees appeared to be intermediate between grandparental ecotypes (Figure 3.4).

The greatest number of F_2 family trees to flower within a flowering year occurred at Boyer in 2010 (106 F_2's flowered) and at Geeveston in 2009 (80 F_2's flowered). Data from these two years were combined for QTL analysis of median flowering time (MFT). Despite there being substantial variation in MFT between sites, a significant QTL on LG8 (129.5 cM) was detected for MFT in both individual-site analyses. Significant QTL by site interactions were detected for this QTL and three other QTL detected for MFT (Table 3.7). Together with the QTL by site interaction detected for one F 10/11 QTL, it appears that flowering in this pedigree is considerably influenced by the environment. A putative QTL for MFT and flowering precocity (F 10/11) co-located on LG4; all other MFT and F 10/11 detected QTL were independent (Figure 3.2 and Table 3.7).

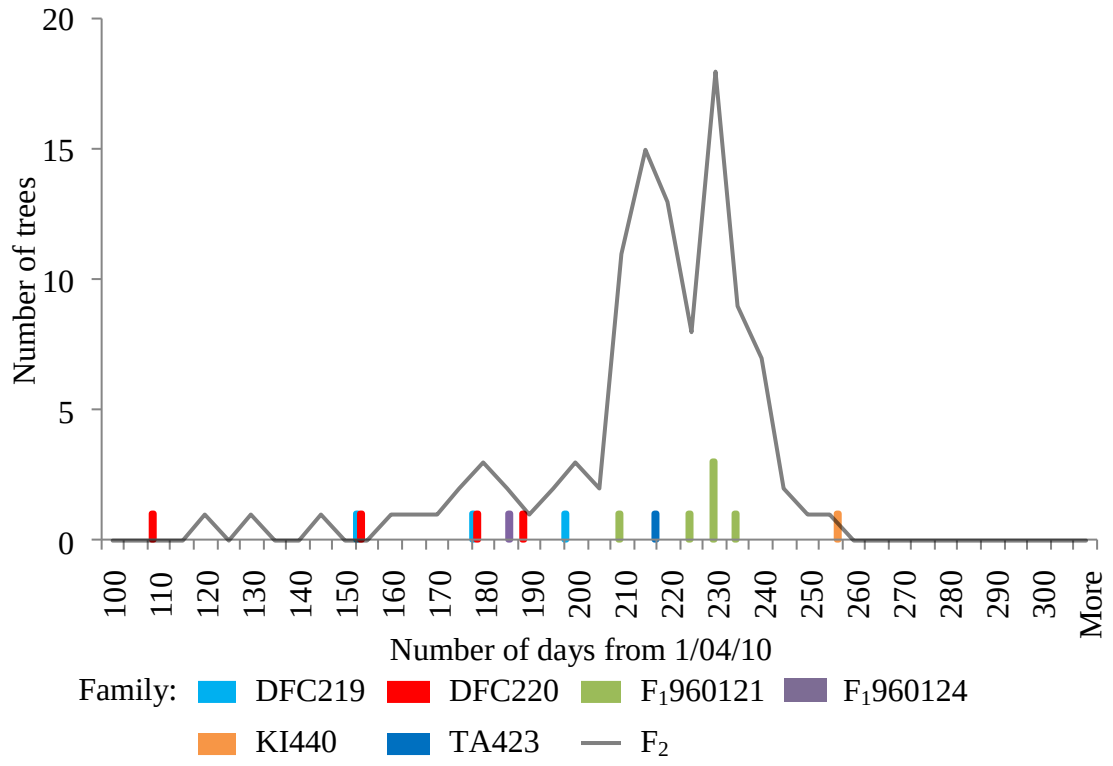

Figure 3.4 Median flowering time (MFT) of flowering trees at Boyer in the 2010 flowering year. Horizontal axis is number of days from 1/4/10; day 110 is 20/7/10 and day 230 is 17/11/10. Bars indicate F_1 and grandparent family trees, the grey line represents F_2 individuals (106 trees).

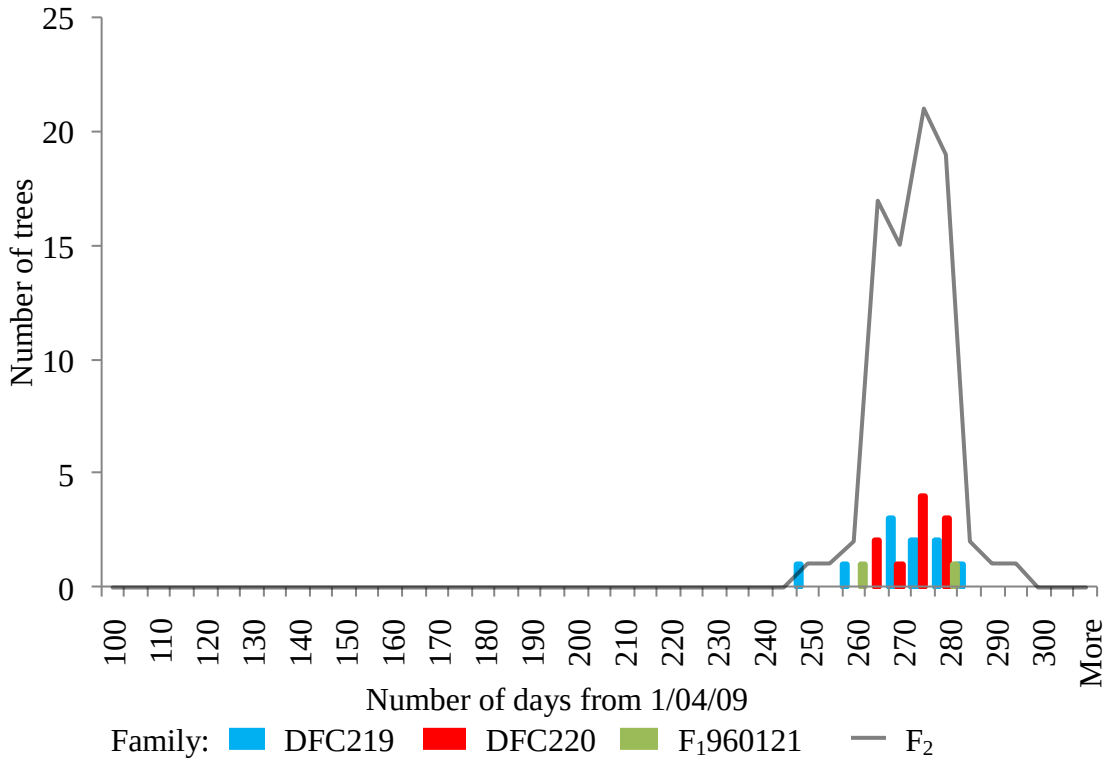

Figure 3.5 Median flowering time (MFT) for all flowering trees at Geeveston in the 2009 flowering year. Horizontal axis is number of days from 1/4/09; day 250 is 7/12/09. Bars indicate F_1 and grandparent family trees, the grey line represents F_2 individuals (80 trees).

3.4 Discussion

The objective of this study was to assess the genetic architecture underlying the phenotypic differences between dwarf and tall ecotypes of the forest tree *E. globulus*, in particular focussing on vegetative and reproductive phase-change traits. The grandparental ecotypes used to generate the large F_2 family of this study had previously been shown to differ in many quantitative traits at either the race or population level (Dutkowski and Potts 1999; Jordan *et al.* 1999; Jordan *et al.* 2000). Variation has also been reported between the Taranna and King Island tall ecotype provenances for height to vegetative phase-change (Freeman 2006) and growth (Volker *et al*. 2008; Freeman *et al.* 2009). In this study, King Island individuals underwent vegetative phase-change and produced adult foliage at a significantly greater number of nodes relative to Taranna individuals (Table 3.5). Therefore, the experimental design employed in this study provided insights into the genetic differentiation between the most extreme dwarf phenotype known in *E. globulus* and two quite diverse tall ecotype families. Using pooled data across grandparental families, the dwarf and tall ecotype families were found to be significantly differentiated in all field-measured traits (contrast tests, all traits $P < 0.0001$; Table 3.5) in a direction which was consistent with the known population level differentiation. These trait differences suggested that numerous traits would be segregating in the F_2 family, thus providing the potential for QTL discovery.

Numerous studies have reported that *E. globulus* individuals from the Wilsons Promontory population are the most precocious of all *E. globulus* populations in their timing to both vegetative and reproductive phase-change (Chambers *et al.* 1997; Dutkowski and Potts 1999; Jordan *et al.* 2000). In this study, Wilsons Promontory grandparental family individuals underwent a significantly faster transition to the adult vegetative state and sexual reproduction (flowering) in contrast to tall ecotype grandparental individuals (Table 3.5). Hamilton *et al.* (2011) recently reported that the timing to vegetative phase-change is genetically variable between *E. globulus* subraces and that this trait is under strong genetic control. As previously found in the *E. risdonii – tenuiramis* complex (Wiltshire *et al.* 1998), the timing to vegetative phase-change in *E. globulus* was also reported to be highly stable across different environments, indicating that minimal genotype-by-environment interaction appears to act on this trait (Hamilton *et al.* 2011). In agreement with Hamilton *et al.* (2011), the node at which vegetative phase-change initiation occurred (NJ) in the F_2 family was highly stable across trials.

This occurred despite there being a significant QTL by environment interaction detected for juvenile internode length, and substantial rainfall, temperature and other site characteristic differences between the Boyer and Geeveston sites (see Table 3.2). While many of the QTL detected in this study were of relatively small effect (less than 5% PVE). The major phase-change QTL detected on LG3 explained 62.8% of the phenotypic variation for NJ. The effects of this QTL segregated from both parents indicating that this QTL differentiated the Wilsons Promontory dwarf ecotype from both tall ecotype families. The alleles of this QTL were detected to have an additive effect and strongly influenced the node at which phase-change occurred. For example, trees which inherited alleles from both dwarf ecotype grandparents underwent vegetative and reproductive phase-change at a significantly smaller number of nodes in comparison to trees which contained either one or two alleles derived from either tall grandparent (Table 3.8). Although other QTL were detected for NJ, NA and NB traits, these were of relatively small effect (PVE 1.4 - 3.7%; Table 3.7), and therefore it appears that in this pedigree the transition to vegetative and reproductive phase change is strongly influenced by this major QTL on LG3. The fact that this major QTL appears to tightly regulate the node position at which vegetative phase-change occurs supports the inference of Jordan *et al*. (1999), who stated that the position of vegetative phase change in *E. globulus* is likely regulated by the number of nodes set rather than absolute age or size. This hypothesis is further supported by the fact that no significant pair-wise phenotypic correlations were detected between node of phase-change traits and tree height (Table 3.6).

Due to the relatively narrow 2-LOD confidence interval of the major phase-change QTL detected on LG3 it was feasible to conduct a search for candidate genes in this region. Recent comparative mapping analyses using high-density linkage maps have indicated that the genomes of *E. grandis* and *E. globulus* are highly syntenic and colinear (this thesis Chapter 1; Hudson *et al.* in press). Therefore, the *E. grandis* genome sequence was used to search for co-locating candidate genes. Micro-RNA miRNA156 is known to be a key regulator of vegetative phase change in annual species (Wu and Poethig 2006) and has been recently shown to play a similar regulatory role in perennial species including *E. globulus* (Wang *et al.* 2011). Six homologues of *Arabidopsis miR156* were identified in the *E. grandis* genome sequence and one of these (named EgMIR156C; Poethig *pers comm.*) was found to be positioned within ~ 1 cM of the QTL peak for NJ on LG3 (Figure 3.3). Assuming that the gene order between *E. grandis* and *E. globulus*

is highly conserved in this region, and given the important role of miRNA156 in regulating vegetative phase-change and that no other genes known to be associated with vegetative phase-change were detected in this QTL region, this micro-RNA represents a good candidate gene for the control of vegetative phase-change in this pedigree and may explain a key difference between dwarf and tall ecotypes of *E. globulus*.

The micro-RNA miRNA156 is highly expressed during early shoot development (which suppresses vegetative phase-change) and declines rapidly during the juvenile-to-adult vegetative transition (Yang *et al.* 2011). Therefore, in dwarf *E. globulus* individuals, miRNA156 levels must decline earlier than in later phase-changing trees. In *Arabidopsis*, a family of eight miRNA156 genes exist (Yang *et al.* 2011) and in a recent study it was shown that only some members of the this gene family contribute substantially to total miRNA156 levels and thus play a major role in the control of vegetative phase-change (Yang *et al.* 2011). For example, Yang *et al.* (2011) examined the separate effects of leaf, root and cotyledon ablation (removal of plant tissue parts) on miRNA156 levels in *Arabidopsis*, maize and *Nicotiana benthamiana* and found that an increase in miRNA156 abundance (*i.e.* suppression of the adult vegetative stage) occurred under leaf primordia ablation only. Primary transcripts of four *miR156* genes (*miR156A*, *miR156B*, *miR156C* and *miR156H*) were then measured under leaf ablation treatment conditions in *Arabidopsis* using real-time polymerase chain reaction analysis and it was found that no change occurred in the primary transcript levels of either *miR156B* or *miR156H*. However, a ~2-fold increase occurred in the primary transcript levels of both *miR156A* and *miR156C*, indicating that only these two genes out of the four measured contributed to the overall increase of miRNA156 levels (Yang *et al.* 2011). Therefore, if multiple *miR156* genes exists in *E. globulus*, similarly to that found in *E. grandis* and many other species, and each *miR156* gene differentially contributes to total miRNA156 levels as found in *Arabidopsis* (Yang *et al.* 2011), then a non-functional copy or deletion of one of the main contributing *miR156* genes (e.g. EgMIR156C in this example) could lead to a substantial reduction in miRNA156 levels. Therefore, if a deletion or mutation of this micro-RNA has occurred in populations of the dwarf ecotype, individuals may be only capable of producing reduced levels of miRNA156. Although this may be sufficient to suppress vegetative phase-change for a limited time, it could be expected that an earlier transition to the adult vegetative state would occur in individuals with a non-functional *miR156C* gene in comparison to individuals with a functional copy (*i.e.* tall ecotype population individuals).

MiRNA156 has also been implicated in the process of flower initiation, and is the only gene known to be involved in both vegetative phase-change and flower initiation (Poethig *pers comm*.; Chen *et al.* 2010). The co-location of QTL for vegetative phase-change traits (NJ and NA) and flower initiation (NB) at the same genomic region as the gene for miRNA156, as well as the highly significantly pair-wise phenotypic correlations between these traits (Table 3.6), is consistent with a pleiotropic effect of miRNA156 on both of these traits. The site stability of NJ and significant environment by trait interactions detected for both NA and NB would account for the reduced significance of QTL for NA and NB. Collectively, these findings support the hypothesised role of miRNA156 in primarily initiating the juvenile-to-adult transition (NJ) and having a declining 'down-stream' effect on the node at which, (1) fully developed adult foliage (NA), and (2) reproductive competence (NB) occurs, respectively.

Although F_2 individuals which inherited both Embra1656 alleles derived from the two dwarf ecotype grandparents underwent NJ, NA and NB transition at significantly fewer nodes in comparison to individuals that contained at least one tall ecotype grandparent allele, the proportion of trees which flowered in each of the four Embra1656 genotype classes were approximately equal (~ 50%). Therefore, although plants (particularly those which inherited both dwarf grandparent Embra1656) may have reached a 'reproductively competent' developmental age (in terms of the number of nodes produced) flowering did not necessary occur. This may have been due to the absence of environmental stimuli that are necessary to induce flowering (Bergonzi and Albani 2011; Simpson *et al.* 1999) or due to the preferential allocation of resources to other traits by these plants; such as growth or survivorship-related traits (e.g. herbivory defence; Obeso 2002). It will be interesting to assess whether Embra1656 genotype class means predict the node of first flower bud initiation in those trees which had not produced flower buds by the end of the 2010 flowering year.

All five flowering precocity (F 10/11) QTL detected in this study were independent of QTL detected for NJ and NB, indicating that genes which initiate the 'switch' to flowering are different to those which control the node at which reproductive phase-change (and also vegetative phase-change) occurs. A previous study which investigated the effect of paclobutrazol (a growth retardant) application on precocious flowering in Wilsons Promontory dwarf ecotype seedlings (Hasan and Reid 1995) supports this finding. It was found that paclobutrazol application significantly increased the

percentage of seedlings that produced flower buds in their first reproductive season (high paclobutrazol treatment; > 95% plants produced buds) in comparison to control plants (~ 30% plants produced buds; Hasan and Reid 1995). However, despite treated seedlings exhibiting an accelerated transition to reproduction, the node number at which flowering occurred was not affected, with no significant difference in the node to first flowering being detected between treated and control plants (Hasan and Reid 1995). In the same study Hasan and Reid (1995) found that paclobutrazol application also stimulated flower bud production in the juvenile vegetative phase of non-precocious *E. globulus* seedlings, supporting the hypothesis that the time to first flowering and vegetative phase-change are genetically independent in *E. globulus* (e.g. Jordan *et al.* 1999). The genetic independence of these developmental traits is necessary to enable a change in their order to occur during plant ontogeny and for the evolution of neotenous species; in which plants become reproductively mature in the juvenile vegetative state (McKinney and McNamara 1991; Zotz *et al.* 2011). This mode of heterochronic evolution has been important within *Eucalyptus* (Potts and Wiltshire 1997), and while it is not known how many genetic changes are required to cause these changes, it has been reported that dramatic heterochronic changes can occur with only a few genetic changes (Wiltshire *et al.* 1994). Although no F_2 individuals of this study were observed to flower in the juvenile vegetative state, two individuals did produce flower buds within intermediate foliage indicating that reproduction can commence prior to trees producing fully adult vegetation (data not shown). These anomalies, taken together with the finding that QTL detected for phase-change and flowering precocity did not co-locate, does provide some evidence for the genetic independence of these processes. However, much stronger evidence was detected in this pedigree to support the rapid evolution of trait acceleration, a form of heterochronic evolution which results in the earlier expression of adult traits during ontogeny. The hypothesised regulatory role of miRNA156c in this family suggests that it may only require a mutation in this, or possibly in other genes of this micro-RNA gene family, to result in this mode of heterochronic evolution.

In addition to the large progeny size of the F_2 family examined, a second major strength of this studies experimental design was the planting of F_2 individuals in two sites. As variable biotic and abiotic environmental factors can modify QTL expression (Brown *et al.* 2003), the detection of QTL at the same (or similar) map positions across multiple sites can substantiate the biological basis of marker-trait associations (Brown *et al.* 2003;

Wheeler *et al.* 2005). However, the comparison of QTL site stability is not straight-forward and was confounded in this study by differences in trial age and sample size; with the smaller Boyer trial (158 F_2's) having only approximately half the number of F_2 individuals as the Geeveston trial. Despite these differences a similar number of QTL were detected in both individual-site analyses. However, only six QTL were detected in both individual-site analyses with five of these being detected in vegetative and reproductive phase-change traits. Although phase-change traits appear to be more site-stable relative to the other trait classes examined in this study, three of the site-stable phase-change QTL included the highly significant NJ, NA and NB QTL which co-located on LG3. Therefore, aside from this locus only a few minor effect phase-change QTL were stable across sites. All five QTL detected for flowering precocity (F 10/11) were detected in individual-site analyses at Geeveston only, and together with the significant QTL by environment interaction detected for 4 out of 10 flowering precocity and median flowering time QTL, flowering traits appeared to be strongly influenced by the environment in this study. Significant genotype by environment (G x E) interactions for flowering traits have been reported in numerous plant studies. For example, Bundock *et al*. (2008) detected variation in flowering precocity between full-sib intra-provenance King Island *E. globulus* families planted in five trials throughout Australia. Significant flowering trait G x E interactions have also been detected in herbaceous species grown in multiple environments (Anderson *et al.* 2011; Brachi *et al.* 2010). Furthermore, G x E interactions have also been argued to explain the differential expression of flowering QTL in different years in almond (Asíns *et al.* 1994), thus providing further evidence that different loci or patterns of gene expression are activated under contrasting environmental conditions for flowering traits.

Although only a few grandparental trees flowered and it was not possible to test for differences between dwarf and tall grandparental families for MFT variation, there was a weak trend for dwarf ecotype grandparental trees to flower earlier in the flowering time period relative to tall ecotype grandparental family trees. Variation in flowering time period has been reported between widespread *E. globulus* populations (Gore and Potts 1995) and therefore it was likely that there was variation between grandparental ecotype families and that this variation would likely segregate in the F_2 family. Despite the apparent G x E interaction observed for flowering traits in this study, a significant QTL was detected on LG8 for MFT in both individual-site analyses. This QTL was detected in different years which had highly contrasting flowering time periods (*i.e.*

Figure 3.4 *vs.* Figure 3.5) suggesting that it may play an important regulatory role in controlling the annual flowering time period in this pedigree. Although many flowering genes are known and the flowering pathway has been well characterised in *Arabidopsis* (Komeda 2004), much less is known about the genetic control of flowering time variation (Ehrenreich and Purugganan 2006). As flowering time difference between populations can restrict gene flow and lead to reproductive isolation, and that these mechanisms have been shown to importantly contribute to local adaptation and ecotypic variation in numerous plant species (Bennington and McGraw 1995; Hall and Willis 2006; Verhoeven *et al.* 2008), this site-stable flowering time QTL detected for MFT may warrant further investigation.

In contrast to the majority of flowering, vegetative and reproductive phase-change QTL which were significant at the genome-wide threshold, QTL detected for tree height and tree-shape traits were generally of lower significance. This was despite tall ecotype grandparental families being significantly differentiated from dwarf ecotype grandparental families in both tree height (mean 4.70 *vs.* 2.30 m; $P < 0.0001$) and apical dominance (H/WP average 2.3 *vs.* 1.7; $P < 0.0001$). Previous eucalypt studies have typically detected only small (e.g. 2-3) numbers of QTL for growth traits (e.g. Bundock *et al.* 2008; Freeman *et al.* 2009; Thumma *et al.* 2010) which has been suggested to reflect the low heritability and polygenic control of growth traits in forest trees (Grattapaglia *et al.* 2009; Hamilton and Potts 2008; Verhaegen *et al.* 1997). Therefore, it could be expected that fewer QTL with lower significance would be detected for growth traits relative to more highly heritable vegetative phase-change (Dutkowski and Potts 1999) and flowering traits (Chambers *et al.* 1997; Gore and Potts 1995) as found in this study. Significant G x E interaction for growth traits have also been detected in eucalypts (e.g. Thumma *et al.* 2010). Therefore, site affects in addition to low heritability and the polygenic control of growth traits may have affected the ability to detect growth trait QTL in both combined and individual-site analyses in this study. A number of forest tree studies have also shown that growth trait QTL are unstable across time (*i.e.* repeat measurements; Kaya *et al.* 1999; Thumma *et al.* 2010; Verhaegen *et al.* 1997), and in some instances QTL have only been detected in later-year measurements. Therefore, if QTL effects associated with either the dwarf or tall ecotype phenotypes have a greater effect in the F_2 family at a later age it could be expected that more significant QTL for growth traits may be detected in this family in the future.

A number of previous QTL studies in *Eucalyptus* have mapped QTL or genes associated with vegetative phase-change and flowering, including a precocious flowering locus in *E. grandis* (*Eef1*; Missiagia *et al.* 2005), homologues of *Arabidopsis* floral genes *AGAMOUS* (*AGE1, AGE2*), *APETALA 1* (*EAP1*) and *LEAFY* homologues (*ELF1*) in *E. globulus* (Thamarus *et al.* 2002), flowering genes *ELF1* and *ELF2* in *E. nitens* (Thumma *et al.* 2010), and a height to vegetative-phase change QTL in *E. globulus* (Freeman 2006). Based on comparative analyses with these studies, none of these mapped genes or QTL were found to co-locate with QTL for related traits in this study. However, considering that the number of flowering genes mapped in *Eucalyptus* to date represents only a few of the likely genes involved in the regulation of flowering (over 60 genes have been shown to regulate flowering time in *Arabidopsis*; Ehrenreich and Purugganan 2006), the failure to detect co-locating flowering loci between studies was not surprising. Furthermore, the *E. grandis* family to which the early flowering locus (*Eef1*) was mapped exhibited an extreme form of precociousness (flowering occurred within 60 to 90 days from germination; Missiagia *et al.* 2005). Thus, it could be expected that different genes underlie this early flowering phenotype to that of the F_2 family in this study. The previously reported QTL for vegetative phase-change was mapped to LG1 in an inter-provenance *E. globulus* F_2 family derived from tall ecotype King Island and Taranna individuals (Freeman 2006). In this family, vegetative phase-change occurred at a mean height of 4.2 m (Freeman 2006) which differed markedly to that found in the dwarf x tall ecotype F_2 family of this study (H-NJ mean 0.78 m; data not shown). Thus, these two mapping pedigrees are highly differentiated and such genetic differentiation is a known factor in contributing to the failure to detect co-locating QTL between studies in forest trees (Sewell and Neale 2000). Although a suggestive QTL was detected in this study for NB on LG1, this QTL was positioned ~50 cM from the vegetative phase-change QTL of Freeman (2006).

For each trait examined in this study, two to five genomic regions affected trait variation and this number of QTL agrees with that found for traits in other ecotypic plant studies. For example, Bratteler *et al*. (2006) detected 2 to 9 QTL for 12 traits which differentiated bladder campion (*Silene vulgaris*) ecotypes and in their meta-analysis which examined the number of QTL detected in 55 traits from six intraspecific crosses, it was found that on average ~ 4 QTL were detected for each morphological trait (Bratteler *et al.* 2006). Apart from two QTL detected for juvenile internode length (Juv-Int-L, LG4 59.6 cM) and NA (LG2 47.8 cM) traits which explained only 3.8 and

2.0 PVE, respectively, all QTL detected in combined-site analyses at the genome-wide significance level were also detected in SML analyses. In a previous study which compared the number of QTL detected in nine barley traits using both SML and composite interval mapping (CIM; a similar method to the rMQM method used in this study; Van Ooijen 2006), Bedo *et al.* (2008) found that although both methods produced similar genome profiles of variance explained (*i.e.* QTL effects), SML detected fewer significant QTL in comparison to the CIM method. This was suggested to reflect the bootstrap-validation step in SML analysis which is presumed to eliminate spuriously detected QTL (Bedo *et al.* 2008). Considering that SML and rMQM mapping are conceptually quite different QTL methods, high confidence can be placed in the majority of QTL reported in this study which were detected using both methods.

This study provided a genome-wide assessment of the genetic differentiation between dwarf and tall ecotypes of the forest tree *E. globulus* for vegetative and reproductive phase-change and growth traits. A total of 22 significant and 11 suggestive QTL were detected (representing 20 unique map positions) across all 11 chromosomes indicating that genetic differentiation between these ecotypes has occurred on a genome-wide scale. However, given that the dwarf and tall ecotypes are likely to differ in many more adaptive traits than what was examined in this study it is probable that many more genetic differences underlie these diverse ecotypes. For example, the major stresses associated with the coastal environments in which dwarf populations inhabit are believed to be related to salt and high wind exposure and or water stress (Chalmers 1992; Foster *et al.* 2007; Jordan *et al.* 2000). These factors can have similar physiological effects in plants by limiting water uptake and subsequently negatively impacting growth (Munns 2002). In these environments it may be beneficial for individuals to produce thicker and/or higher density leaves in comparison to individuals in more productive habitats. Such adaptations can increase leaf water-use efficiency, render leaves less susceptible to physical hazards (*i.e.* wind) and increase the life-span of leaves which may be preferential in environments where slow tissue turnover is favoured (reviewed by Reich *et al.* 2003; Wright *et al.* 2001). Therefore, it could be expected that the dwarf and tall ecotype families in this study may differ in many leaf characteristics, similarly to that reported in Dutkowski and Potts (1999) between Wilsons Promontory dwarf ecotype and tall ecotype provenance individuals of *E. globulus*. As all F_2 individuals have now produced near-full canopies of adult foliage,

morphological and physiological leaf traits could now be assessed in order to further characterise the differences between these ecotypes.

Molecular genetic evidence indicates that the three naturally occurring dwarf ecotype populations of *E. globulus* which occur at Wilsons Promontory in Victoria and in Tasmania at Cape Tourville and Maria Island have co-evolved independently from neighbouring tall *E. globulus* populations at each location (Foster *et al.* 2007). Although the Wilsons Promontory population is located some distance from its closest extant neighbouring tall population (Tidal River, ~13 km), dwarf and tall populations occur within close proximity (~300 m) at Cape Tourville and strong clinal variation in growth, the timing to vegetative phase-change and first flowering have been shown to occur between these population (Jordan *et al.* 2000). While in dwarf ecotype individuals, early vegetative phase-change is associated with slow growth and low height to phase-change, field trial experiments representing most tall ecotype *E. globulus* provenances have shown opposing trends, with early vegetative phase-change being associated with high height of phase-change and fast growth (Jordan *et al.* 2000). These differences, in addition to the genetic independence of vegetative and reproductive phase-change (Jordan *et al.* 1999) and independence of first flowering and growth (Chambers *et al.* 1997) argues that selection pressures associated with the coastal habitat have acted independently, and in parallel, on vegetative phase-change, flowering and growth traits during the divergence of dwarf ecotypes from neighbouring tall populations (Jordan *et al.* 2000). Although only a few minor QTL were detected for growth traits in this study, the major QTL detected on LG3 and hypothesised role of EgMIR156C in regulating vegetative phase-change in this family may represent a key target of selection during the evolutionary divergence of dwarf individuals from neighbouring tall *E. globulus* populations. In fact, the ability of *E. globulus* individuals to colonise the exposed coastal habitat in which dwarf ecotypes occur may have been dependent on the development of early phase-change, as this would have provided trees with physically stronger adult foliage; thus being less susceptible to wind damage earlier in ontology. The hypothesised involvement of EgMIR156C in phase-change in this study strengthens the recently shown involvement of miR156 in *E. globulus* vegetative phase-change regulation (Wang *et al.* 2011) and contributes to the increasing body of research focussing on understanding the molecular genetics of vegetative phase-change due to the importance of this process in plant development, adaptation and speciation processes. Experiments are currently being designed to clone, characterise and assess

the functional role of this candidate in this pedigree. It will then be possible to conduct population-level tests for associations between sequence polymorphisms of EgMIR156C with phase-change trait values between the naturally occurring dwarf and tall neighbouring populations. These future studies offer the potential to confirm the role of miRNA156 and its prevalence in the evolution of dwarf ecotype *E. globulus* populations.

Chapter 4 A 'Genomic Atlas' for *Eucalyptus*: population genomic studies in five subgenus *Symphyomyrtus* species

4.1 Introduction

Adaptive evolution is the process by which populations genetically diverge in response to natural selection. At the molecular level, selection shapes patterns of genetic variation between individuals, populations, and species, and can do so differentially across genomes (Hohenlohe *et al.* 2010). The identification of genomic regions which are highly differentiated between populations and species can provide insights into the genetic mechanisms involved in adaptive phenotypic differentiation and speciation (Bodenes *et al.* 1997; Harr 2006; Johansson *et al.* 2010). Population genomics is one approach that can be used to identify differentiated regions in genomes (Stinchcombe and Hoekstra 2007). This method relies on the concept that selection distorts patterns of neutral variation throughout the genome in predictable ways and that these patterns can be detected using a large number of molecular markers distributed throughout a species genome (Pritchard *et al.* 2010). For example, a standard neutral model predicts that genetic drift and migration affects all loci equally, whereas mutation and natural selection affects specific loci and results in detectable signatures of selection (Achere *et al.* 2005). This principle can be illustrated by a selective sweep, in which selection favours an allele resulting in the allele being driven to fixation within a population. In this situation, selection for the favoured allele leads to a reduction in genetic diversity at the selected locus, and in the immediate genomic area due to genetic hitch-hiking, relative to regions not under selection (Savolainen and Pyhäjärvi 2007; Stinchcombe and Hoekstra 2007). When multiple populations (or species) are considered, variable selection pressures, for example associated with different habitats, may select for the different alleles of a gene. This may result in selective sweeps occurring in each population, but in opposite directions (*i.e.* disruptive selection). Under this scenario, selection not only reduces within population (or species) diversity at the selected locus, but increases population divergence at the selected locus (Scotti-Saintagne *et al.* 2004). The genes underlying the adaptive trait(s) being selected for, or loci linked to them, can be identified using outlier tests, which aim to identify loci having higher measures of divergence (e.g. marker F_{ST}) relative to the neutral distribution of divergence estimated using many hundreds or thousands of genetic markers (Butlin 2010). Once these outlier markers have been detected, their genome position can be identified using either, (1)

genetic linkage maps, or for a limited number of organisms, (2) a genome reference sequence. This then allows for the organisation and distribution of loci under putative selection to be examined (Bodenes *et al.* 1997).

Genome-wide analyses depend on the availability of large genetic resources [e.g. reference genome sequences, single nucleotide polymorphism (SNP) genotyping chips, pedigree information, inbred strains etc.; Hohenlohe *et al.* 2010] and were initially restricted to model species including *Drosophila* and humans (Hohenlohe *et al.* 2010). The first high-density, genome-wide population genomic analysis was conducted by Akey *et al.* (2002). This study calculated individual marker F_{ST} values among three human populations (African-American, east Asian and European-American) for more than 26,000 SNP markers and examined the distribution of marker F_{ST} values at genome, chromosome and gene levels. One hundred and seventy-four candidate genes were identified as being subject to selection and the mapping of these markers provided a framework for the first natural selection map of the human genome (Akey *et al.* 2002). Scotti-Saintagne *et al.* (2004) extended the concept of population genomics to the species-level in their investigation of differentiation between European oak species (*Quercus robur* and *Q. petraea*). Oak populations of both species were sampled throughout their natural distribution with data grouped over populations to obtain allele frequencies for each species. Outlier analysis detected 47 markers with significant G_{ST} deviations (analogous to F_{ST}) from neutral expectation and 20 of these markers could be mapped on the oak genome using linkage map information. Despite only mapping half of the outlier loci detected, and these loci being distributed across 9 out of 12 linkage groups, a few 'hot-spots' containing tightly linked outlier loci were identified (Scotti-Saintagne *et al.* 2004). Once these 'hot-spot' regions are identified, the species (or a closely related species) genome sequence can be used to identify candidate genes in these regions for use in further functional studies (Butlin 2010; Hohenlohe *et al.* 2010). Such an approach has provided insights into the process of speciation between house mice species. For example, Harr (2006) identified several regions which differentiated inbred laboratory lines of two house mice subspecies (*Mus musculus musculus* and *M. m. domesticus*). Analyses were then conducted using individuals from wild populations which identified differentiation in these same genomic regions. Having replicated their study and verified these differentiated regions, an examination of the genes residing in these areas was undertaken and it was found that genes associated with immunity-response and olfaction were highly over-represented relative to other gene classes (Harr

2006). Further evidence for these gene classes playing a role in speciation has been provided by studies investigating human speciation, which have identified these gene classes as being under directional selection (see Nielsen *et al.* 2005).

Recent advances in sequencing technology and the development of various array-based genotyping techniques are now providing the necessary genomic resources required for genome-wide studies in many non-model species (Hohenlohe *et al.* 2010). For example, in the forest tree genus *Eucalyptus,* the release of an assembled genome sequence for *E. grandis* (www.phytozome.net/) and development of a eucalypt Diversity Arrays Technology (DArT) microarray genotyping system (Sansaloni *et al.* 2010; Steane *et al.* 2011) have dramatically increased the feasibility of conducting genome-wide studies. DArT is a high-throughput, genome-wide genotyping system which can provide several thousand polymorphic markers per study. A composite genetic linkage map containing 4135 markers (including 3909 DArT markers; this thesis Chapter 2) was recently produced for subgenus *Symphyomyrtus* species. This map provides a valuable high-density reference map for genome-wide studies in these species.

The adaptability, growth and wood qualities of eucalypts have lead to eucalypt species and their hybrids becoming the most planted hardwood tree species world-wide (Doughty 2000). There are approximately 20 million hectares of eucalypt plantations currently established for solid timber, pulp and paper production (Iglesias-Trabado and Wilstermann 2008). Many of the most commercially important species belong to subgenus *Symphyomyrtus* sections *Latoangulatae*, *Maidenaria* and *Exsertaria* (Grattapaglia and Kirst 2008). For example, *E. grandis* and *E. urophylla* (section *Latoangulatae*) are commonly used to produce fast-growing, disease tolerant hybrids for tropical and subtropical plantation forestry (Grattapaglia and Kirst 2008; Myburg *et al.* 2007). In temperate regions, *Eucalyptus globulus* spp. *globulus* (hereafter referred to as *E. globulus*) and *E. nitens* (section *Maidenaria*) are the preferred pulpwood plantation species (Dutkowski *et al.* 2001; Grattapaglia and Kirst 2008) and from section *Exsertaria, E. camaldulensis* is an important species for the production of construction timber and paper pulp across a wide range of arid, temperate and tropical countries (Butcher *et al.* 2009; McDonald *et al.* 2009). Additionally, *E. camaldulensis* is widely planted in northern Africa and in the Indian sub-continent as a source of fuel and has been used in many land rehabilitation projects; particularly for addressing soil salinity issues (Butcher *et al.* 2009; Marcar and Crawford 2004). These five species are

probably the most important eucalypts used in plantations throughout the world (Harwood *pers comm.*).

The natural distributions of these species extend across an extraordinary range of habitats. For example, *E. globulus* is predominantly found in coastal, temperate forests in Tasmania (~ 40-43°S; Williams and Potts 1996). *Eucalyptus nitens* usually grows at an altitude of 1000 m or above and is distributed along the Great Dividing Range in eastern Australia from the Central Highlands of Victoria (~ 37°S) to northern New South Wales (~ 30°S; Cook and Ladiges 1991). Further north along the eastern seaboard of Australia, *E. grandis* is found to have a fairly continuous distribution in coastal and sub-coastal regions from the mid-north coast of NSW (33°S) to south-east Queensland (QLD, 25°S). Smaller *E. grandis* populations are also found further north on the ranges inland of Townsville in northern Queensland (16-19°S; Brooker and Kleinig 2004; Jones *et al.* 2006). *Eucalyptus camaldulensis* has the most widespread distribution of any eucalypt, ranging from ~12 to 37°S latitude and occurring in all mainland Australian states (McDonald *et al.* 2009). Further north, and being one of the few eucalypts found exclusively outside of Australia, *E. urophylla* occurs on a series of islands in the Sunda archipelago in eastern Indonesia (~ 8-10°S; Payn *et al.* 2008). This species has the greatest altitudinal range of any eucalypt species, occurring from close to sea level (70 m) up to almost 3,000 m elevation (Gunn and McDonald 1991). Genome-wide studies thus provide an opportunity to examine the molecular basis of adaptation between these ecologically diverse and economically important species.

In this study, genome-wide analyses were conducted using range-wide samples (~90 each species) of five eucalypt species; *E. grandis*, *E. urophylla*, *E. globulus*, *E. nitens* and *E. camaldulensis*. Samples of each species were genotyped with DArT which provided 2207 common markers across species. Genetic diversity estimates were then obtained for each marker in each species and in order to investigate the molecular basis of species differentiation outlier tests were performed for each of ten pair-wise species comparisons. Results were then examined at the genome-level by plotting species genetic diversity estimates and outlier markers on the reference *Eucalyptus* multi-species composite linkage map (this thesis Chapter 2). This was essentially a pilot study, with the specific aims being, (1) to assess the ability of DArT markers in combination with current genomic resources to conduct genome-wide analyses in subgenus *Symphyomyrtus* species, (2) explore the possibility of developing a so-called 'Genomic Atlas' for subgenus *Symphyomyrtus* species. The Genomic Atlas would be an inventory

of genetic diversity estimates for multiple eucalypt species mapped on the 11 chromosomes of *Eucalyptus*.

4.2 Materials and Methods

4.2.1 Sampling methods

Five species of subgenus *Symphyomyrtus* were used for this study; *E. globulus, E. nitens,* (section *Maidenaria*), *E. camaldulensis, E. grandis* (section *Latoangulatae*) *and E. urophylla* (section *Exsertaria*). Each species was represented by 84 to 93 samples from across their range-wide distributions. *Eucalyptus globulus* material was obtained from the large plant tissue collection (field collected samples stored at -80°C) held in the School of Plant Science at the University of Tasmania (UTAS). These 84 samples represented the 13 geographical races of *E. globulus* (Dutkowski and Potts 1999). *Eucalyptus nitens* leaf tissue was sourced from two forestry research trials in Tasmania; at Hollowtree (Norske Skog/Forestry Tasmania) and West Ridgly (Gunns Ltd.). Both trials were established from native forest open-pollinated seed collections. The majority of *E. nitens* samples included in this study were from central Victorian populations (65 out of 85 samples); which is an important source of genetic material for plantation and breeding programs world-wide (Dutkowski *et al.* 2001). Four to six samples from each of four small, disjunct *E. nitens* populations (Byrne *et al.* 1998; Cook and Ladiges 1991) located in northern and southern NSW were also included. *Eucalyptus nitens* and *E. globulus* genomic DNA was isolated from leaf tissue at UTAS using a modified CTAB extraction process (Doyle and Doyle 1990).

Isolated DNA for *E. camaldulensis* samples were obtained from CSIRO Plant Industries in Canberra, ACT (CSIRO Black Mountain Laboratories). These samples had been collected as part of a range-wide study (Butcher *et al.* 2009). Ninety-two samples from 88 populations were obtained; covering the species distribution on mainland Australia and including all seven recognised subspecies (McDonald *et al.* 2009). Isolated DNA for *E. grandis* and *E. urophylla* samples were supplied by the University of Pretoria (South Africa). DNA for these samples was extracted from field trial grown trees in South Africa (Sappi and Mondi Business Paper forestry companies) established from natural stand open-pollinated seed collections. *Eucalyptus urophylla* material originated from Camcore seed collections (see Pepe *et al.* 2004).

For *E. grandis,* nineteen (Coffs Harbour region) to twenty-six (southern NSW region; Table 4.1) samples from each of four genetically differentiated geographic regions (Jones *et al.* 2006) were included in this study; resulting in 90 samples being assessed.

Ninety-three *E. urophylla* samples representing four (Alor) to eleven (Timor; Table S4.1) provenances from each of the seven Lesser Sunda islands on which the species occurs were included in this study. *Eucalyptus urophylla* samples were not available from East Timor populations; due to political unrest at the time of Camcore seed collections (Payn *et al.* 2008; Pepe *et al.* 2004). The altitudinal range of *E. urophylla* samples included in this study ranged from 385 (Alor) to 1900 m (Timor; data not shown). For *E. nitens*, *E. grandis* and *E. urophylla* samples collected in field trials, only one progeny per family was used. Sample locality information is summarised in Table 4.1; more detailed information is available in Table S4.1.

Table 4.1 Locality information and number of samples for each of five *Eucalyptus* species.

Species, population / race / subspecies[a]	State[b]	*n*
E. camaldulensis		
acuta (northern)	QLD	4
acuta (southern)	QLD	4
camaldulensis	QLD/NSW/SA/VIC	22
minima	SA	5
obtusa	NT/QLD/WA	24
refulgens	WA	14
simulata	QLD	2
subcinerea	NSW/NT/QLD/SA/WA	17
Subtotal		92
E. globulus		
Northeast Tasmania	TAS	12
Southeast Tasmania	TAS	8
Southern Tasmania	TAS	9
Dromedary	TAS	2
Recherche Bay	TAS	2
Western Tasmania	TAS	9
Furneaux group	TAS	8
King Island	TAS	8
Eastern Otways	VIC	5
Western Otways	VIC	7
Southern Gippsland	VIC	8
Strzelecki Ranges	VIC	2
Wilson Promontory	VIC	4
Subtotal		84
E. nitens		
Ebor	NSW	5

Barrington Tops	NSW	4
Tallaganda	NSW	6
Nimitabel	NSW	5
Connors Plains	VIC	5
Mt Wellington	VIC	1
Central Victoria - North	VIC	22
Central Victoria - South	VIC	37
Subtotal		85
E. grandis		
Atherton	QLD	25
Gympie	QLD	20
Coffs Harbour	NSW	19
Southern NSW	NSW	26
Subtotal		90
E. urophylla		
Adonara	N/A	14
Alor	N/A	9
Flores	N/A	13
Lomblen (Lembata)	N/A	14
Pantar	N/A	10
Timor	N/A	19
Wetar	N/A	14
Subtotal		93
Total		444

[a]Population / race / subspecies; *E. camaldulensis* subspecies follow McDonald *et al.* (2009); *E. globulus* race follows Dutkowski and Potts (1999); *E. nitens* - population for NSW samples, Central Victorian samples classified according to Dutkowski *et al* (2001); *E. grandis* – geographical region; and *E. urophylla* - Lesser Sunda Island of eastern Indonesia. [b]Australian state of sample; not applicable (N/A) for *E. urophylla*.

4.2.2 DArT genotyping

Samples were genotyped by DArT Pty Ltd (Canberra, ACT). Sample DNA (15 µL at 50-70 ng/µL) was arranged in five 96-well plates. Each plate contained five species blocks; one block per species containing 18-19 samples. Blocks were randomly arranged within plates with samples randomly distributed within blocks. Twenty repeat samples were included to assess genotyping error rates. In brief, the DArT genotyping microarray used for sample assay contained 7680 fragments (printed in duplicate) that were generated from genomic representations generated from 64 eucalypt species using a *Pst*I/*Taq*I complexity reduction method (Sansaloni *et al.* 2010). In sample genotyping, DNA was firstly digested with *Pst*I/*Taq*I restriction enzymes producing a genomic

representation (*i.e.* digested fragments) which was subsequently fluorescently labelled and hybridised to the genotyping microarray. Following hybridisation, microarrays were scanned and marker hybridisation intensities were analysed using DArTsoft (version 7.44) software (http://www.diversityarrays.com/). Marker presence (1) or absence (0) scores were generated for each sample along with quality parameters for each marker as described in Sansaloni *et al.* (2010). All five species were analysed together.

4.2.3 Genetic analyses

4.2.3.1 Preliminary analyses: STRUCTURE and GenAlEx

STRUCTURE 2.3 (Pritchard *et al.* 2000) analyses were performed to examine whether DArT markers could correctly assign individuals to species groups using no prior species classification information. STRUCTURE uses Bayesian model-based clustering to identify genetic structure amongst individuals. Samples were assigned (probabilistically) to populations (*K*) based on their genotype using the admixture model and having a 100,000 burn in period and 100,000 MCMC iterations. Ten independent runs at each *K* value (*K* = 1 to 10) were performed with the best *K* (*i.e.* number of species groups) determined according to Pritchard *et al.* (2000) and or Evanno *et al.* (2005). The online version of Structure Harvester (v0.56.4 May 2010; http://taylor0.biology.ucla.edu/struct_harvest/) was used to view STRUCTURE results. The programs CLUMPP (Jakobsson and Rosenberg 2007) and Distruct (Rosenberg 2004) were used to average member (*K* group) coefficients over repeat runs and to graphically display averages in line segment plots, respectively.

The program GenAlEx (Peakall and Smouse 2006) was used to conduct an analysis of molecular variance (AMOVA). A pair-wise individual genetic distance matrix was calculated following Huff *et al*. (1993) which was then used to estimate pair-wise species Φ*pt* (a measure of population genetic differentiation for binary data that is analogous to F_{ST}; Peakall and Smouse 2006). Significance was determined by 999 permutations of the data.

4.2.3.2 Fixed marker differences

From DArT marker presence (1) and absence (0) genotype scores, marker '1 ratio' scores were calculated for each marker within each species as:

1 ratio = number of 1 scores / (number of samples – number of missing data)

This measure was used to identify fixed marker differences in pair-wise species comparisons by calculating 1 ratio differences.

4.2.3.3 Genetic diversity (*H*)

Single locus and mean (species level) genetic diversity estimates (*H*) were calculated for each species. For dominant markers such as DArT, *H* is estimated using the frequency of the null homozygote (*Q*; e.g. '0' scores) and the species fixation index (Wright's *F* or Hardy-Weinberg deviation; Kremer *et al.* 2005). While *Q* can be obtained directly from genotype scores, *F* estimates can generally only be obtained from population genetic studies using co-dominant markers. Although single locus estimations of *H* are strongly influenced by variations in *F* (range: 0 = Hardy-Weinberg equilibrium, to 1 = no heterozygote class or complete inbreeding), Kremer *et al.* (2005) has shown that species level genetic diversity estimates (mean over all loci) are robust to deviations in *F*. This is due to the negative and positive biases associated with calculating *H* (without incorporating *F*) being averaged out over a large number of markers. Furthermore, this robustness has been shown to be strongest under balanced or 'U-shaped' *Q* frequency profiles (Kremer *et al.* 2005). Therefore, *Q* frequency profiles were inspected for each species and two genetic diversity measures representing the extreme values of *H*, H_{HW} (assuming Hardy-Weinberg equilibrium, $F = 0$) and H_{PH} (assuming no heterozygote, $F = 1$), also known as the phenotypic diversity (Mariette *et al.* 2002), were calculated for each locus within species as:

$$H_{HW} = 2\sqrt{Q}\,(1 - \sqrt{Q})$$

$$H_{PH} = 2Q\,(1 - Q)$$

As these two estimates represent the upper and lower values of *H*, the true estimation of *H* (*i.e.* if it were possible to include an *F* estimate) will be within this range. The non-parametric Kruskal-Wallis test was used to test for significance between species and linkage group (across species) H_{HW} values using the PROC NPAR1WAY WILCOXON statement in SAS (Version 9.2; SAS Institute, Cary, USA).

4.2.3.4 Bayescan outlier marker detection

The program Bayescan V2.01 (Foll and Gaggiotti 2008) was used to identify loci under putative selection in each of 10 pair-wise species comparisons. The basic rationale of the program is that loci under directional selection will show a larger genetic differentiation (marker F_{ST}) than neutral loci, and that loci subject to balancing selection

will show lower genetic differentiation than neutral expectations. Bayescan firstly calculates marker F_{ST} and secondly implements a Bayesian method to identify markers with F_{ST} coefficients significantly different from those expected under neutral theory (Foll and Gaggiotti 2008). In each pair-wise species comparison, DArT markers scored as '1' in < 2% and ≥ 98% of individuals (across both species) were excluded from analyses as this 2-98% criterion is reported to be a more conservative approach for detecting outliers. Default Bayescan V2.01 settings were used which included a prior odds value of 10. Posterior probabilities were calculated for each locus being subject to selection. These probabilities are not directly comparable to standard *P*-values; instead they are interpreted using the so-called 'Bayes factors'. Markers having a posterior probability threshold of ≥ 0.76 are assessed to be under 'substantial' selection, ≥ 0.91 'strong' selection, ≥ 0.97 'very strong' selection and ≥ 0.99 'decisive' selection following Jeffrey's Bayes factor interpretation (Foll and Gaggiotti 2008). In Jeffrey's interpretation, markers having a posterior probability of ≥ 0.50 < 0.76 are reported to be 'barely worth mentioning' in terms of selection. However, as an aim of this study was to identify loci and genomic regions putatively involved in species differentiation, markers with a posterior probability threshold of ≥ 0.50 were of interest and were considered as potential outliers. It is important to consider that in this study, validation of a locus or genomic region under selection could be provided through 'taxonomic replication'. While most population genomic studies conduct either, a single outlier marker detection test, or may have replication across environments within a species, this study conducted 10 pair-wise species comparisons with several being taxonomically independent; e.g. the *E. globulus* vs. *E. nitens* (section *Maidenaria*) comparison was an independent comparison to the section *Latoangulatae E. grandis* vs. *E. urophylla* comparison.

4.2.4 Mapping of genetic measures: The 'Genomic Atlas'

The Genomic Atlas was developed by plotting genetic diversity estimates and outlier markers on the *Eucalyptus* multi-species composite linkage map (Chapter 2 of this thesis; hereafter referred to as the composite map). Eighty-four DArT markers had been mapped to two or more linkage groups in the composite map and were not used for plotting diversity estimates or outlier markers. H_{HW} was used to examine genome-wide genetic diversity. This measure was chosen as eucalypts preferentially out-cross (Potts *et al.* 2008; Potts and Wiltshire 1997), and thus, species *F* estimates are expected to be closer to 0 (Hardy-Weinberg equilibrium; H_{HW}) than to 1 (inbreeding species; H_{PH}). An 11 marker running average was calculated for H_{HW}; for each marker using five flanking

markers on either each side of the marker. A minimum of six markers were used to calculate H_{HW} averages for markers mapped at the end of linkage groups due to flanking markers being present in one direction only. Figures were generated using GraphPad Prism (V5; GraphPad Software, Inc. USA) or PSI-plot (V10; Poly Software International, USA).

4.3 Results

4.3.1 Dataset curation

Data was received for 4706 markers and 469 samples following DArT genotyping. One *E. camaldulensis* sample (C_14_142) was not genotyped due to poor DNA quality. Five samples (C_76_401, GR_DP149, U_DP049, GR_DP169 and GL_529) had greater than 10% missing data (range 10.3-18.3%) and were excluded from analyses. DArT genotyping generates a number of marker quality parameters. The DArT array contains 15360 markers (clones), comprising two randomly positioned clones for each of the 7680 markers (Sansaloni *et al.* 2010). From this, a 'Reproducibility' value is generated as part of DArT quality control (*i.e.* internal repeats); being a percent score of how reproducible a marker score is within a sample hybridisation. Independent to this, 20 repeat samples were included to enable an independent assessment of DArT genotyping 'repeatability'. For each of these 20 repeat samples, DArT marker scores were compared and repeatability was found to be high; with an average marker repeatability of 98.8% (standard error 0.11) for the 4706 markers. Two hundred and ten DArT markers were found to have incongruent marker scores at a frequency of ≥ 0.1 (*i.e.* ≥ 2 out of 20 repeat samples). These markers were considered to be unreliable and were excluded; reducing the dataset to 4496 markers.

Due to the anonymous nature of DArT clones (e.g. markers) during microarray development it is possible for the same clone to be represented more than once on the final operational genotyping array; thus introducing a level of marker redundancy (Sansaloni *et al.* 2010). Most DArT marker clones have been sequenced (GenBank accession numbers HR865291 - HR872186; Kullan *et al.* in press) and a recent sequence-based redundancy analysis (based on 6918 of the 7680 markers contained on the genotyping array; Sansaloni and Petroli unpublished) was used to identify redundant markers within the 4496 marker dataset. Most of the 4496 markers (4147 or 93%) were included in the redundancy analysis of Sansaloni and Petroli (unpublished). Of these, 1241 were found to be unique markers whereas 2906 markers were allocated to 966 bins containing two-or-more markers. From each of these multi-marker bins, the highest quality marker (e.g. least missing data and highest DArT reproducibility score) was retained. This resulted in a further 1940 markers being excluded; 47% of the 4147 markers assessed. Due to the high proportion of marker redundancy within the dataset, the 349 DArT markers which were not included in the redundancy analysis of Sansaloni

and Petroli (unpublished) were also excluded; as it could be expected that approximately half of these markers could also be redundant.

Following the removal of repeat samples (20), the final dataset contained 84 (*E. globulus*) to 93 (*E. urophylla*) samples per species (Table 4.1 and Table S4.1) and totalled 444 samples x 2207 DArT markers. DArT reproducibility scores were ≥ 97% for all 2207 markers with 89% (1972) of markers having ≥ 99% reproducibility. The species origins of these 2207 DArT markers are presented in Table 4.2. For markers with known species origins, most were derived from *E. grandis*, *E. globulus* and *E. urophylla* (Table 4.2). Markers of unknown origin were derived from a pooled collection of DNA taken from multiple species (Sansaloni *et al.* 2010).

Table 4.2 Species origins of DArT markers used in analyses. Number and percentage of total number of DArT markers are shown for the full dataset (2207 markers) and markers mapped (1408) on the *Eucalyptus* multi-species composite linkage map. Markers of 'unknown' origin were derived from DNA which was pooled from multiple species (see Sansaloni *et al.* 2010).

Subgenus, section and species	Total		Mapped	
	n	%[a]	*n*	%[b]
Subgenus *Symphyomyrtus*				
Latoangulatae				
E. grandis	700	31.7	410	29.1
E. urophylla	126	5.7	86	6.1
E. grandis x E. urophylla hybrid	142	6.4	117	8.3
Subtotal	968	43.9	613	43.5
Maidenaria				
E. globulus	382	17.3	255	18.1
E. nitens	110	5.0	69	4.9
Subtotal	492	22.3	324	23.0
Exsertaria				
E. camaldulensis	38	1.7	20	1.4
Subgenus *Eucalyptus*				
Psuedophloia				
E. pilularis	54	2.4	34	2.4
Corymbia species	31	1.4	17	1.2
Known species origins subtotal	1583	71.7	1008	71.6
Unknown species origins	624	28.3	400	28.4
Total	2207	100.0	1408	100.0

[a]Percentage of total 2207 markers. [b]Percentage of 1408 mapped markers.

4.3.2 STRUCTURE and GenAlEx

The program STRUCTURE was used to assign individuals of the 444 sample x 2207 marker dataset to *K* groups and to determine the optimal number of groups. No prior

species classification information was used in this analysis. Consistent with the peak of Evanno's Δ*K* (Evanno et al. 2005) at *K*=5 (Figure 4.1), the genetic structure of the five species dataset was partitioned into five groups (Figure 4.2).

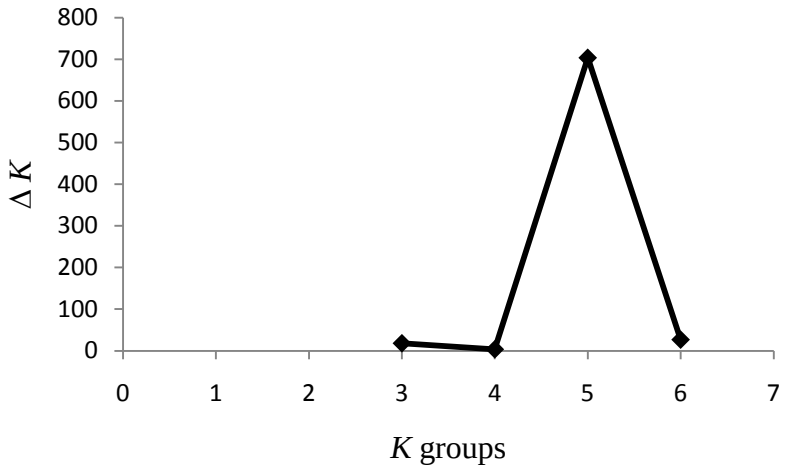

Figure 4.1 Evanno's Δ*K* for five species 444 individual x 2207 marker STRUCTURE analysis; mean calculated from 10 independent runs.

All individuals were correctly assigned with high probability to their respective 'species' *K* group. For all species *K* groups, the mean membership coefficient (*Q*) of individuals assigned to each group was ≥ 0.98. *Eucalyptus nitens* samples (yellow cluster; Figure 4.2) from Ebor (5 individuals; population 17) and Barrington Tops (4 individuals; population 18) populations in northern NSW showed moderate membership to the *E. globulus* species group (blue cluster; Figure 4.2). The mean membership coefficients of these nine samples were 0.87 and 0.11 to *E. nitens* and *E. globulus* species *K* groups, respectively.

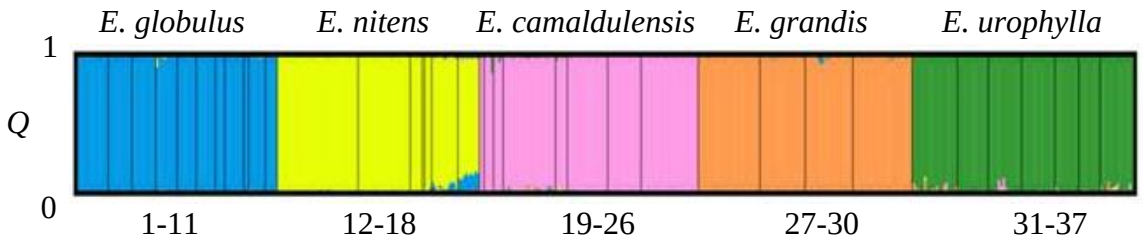

Figure 4.2 STRUCTURE line segment plot showing the mean membership coefficient (*Q*, vertical axis, average of 10 independent runs) for 444 individuals at $K = 5$; analysis conducted on 2207 markers. *K* groups are colour coded with species given above each group. No prior species information was used in analyses. Individuals are represented by a vertical line and are grouped by populations (indicated by black vertical lines) within each species group cluster. Populations are numbered 1-37 from left to right (see Table S4.1).

An analysis of molecular variance (AMOVA) showed molecular variance between species accounted for 40% of the variation which was significant ($P < 0.001$). Most of the molecular variation occurred within species (60%). All species were significantly differentiated ($P < 0.001$) based on pair-wise Φ*pt* values. Differentiation was greatest between *E. nitens* (section *Maidenaria*) and *E. grandis* (section *Latoangulatae*) and was typically highest in inter-section pair-wise species comparisons (e.g. between section *Maidenaria* species *E. globulus* and *E. nitens* with species from other sections; Table 4.3). However, *E. camaldulensis* and *E. urophylla* were less differentiated (inter-section comparison, Φ*pt* 0.283; Table 4.3) compared to species within sections *Latoangulatae* (*E. grandis* and *E. urophylla*; Φ*pt* 0.304) and *Maidenaria* (*E. nitens* and *E. globulus*, Φ*pt* 0.361; Table 4.3).

Table 4.3 Species pair-wise Φ*pt* values (an analogous measure to F_{ST}). All values were significant ($P < 0.001$) based on 999 permutations of the data.

Species[a]	*E. globulus*	*E. nitens*	*E. camaldulensis*	*E. grandis*	*E. urophylla*
E. globulus	0.000				
E. nitens	0.361	0.000			
E. camaldulensis	0.426	0.470	0.000		
E. grandis	0.447	0.495	0.366	0.000	
E. urophylla	0.416	0.461	0.283	0.304	0.000

[a]Section *Maidenaria* - *E. globulus* and *E. nitens*; Section *Latoangulatae* - *E. urophylla* and *E. grandis*; Section *Exsertaria* - *E. camaldulensis.*

4.3.3 Markers mapped on the composite map

Of the 2207 markers, 1448 (66%) could be mapped on the composite linkage map. This included 40 multicopy markers mapped to two or more linkage groups (LGs). These markers were excluded from genetic diversity plots and outlier marker mapping (see below), thereby reducing the number of markers which could be assigned to a single LG to 1408. The mean number of markers mapped on LGs was 128 and ranged from 81 (LG4) to 185 (LG8; Table 4.4).

4.3.4 Genetic diversity estimates

For each species, individual marker H_{HW} and H_{PH} estimates were calculated for all 2207 markers. *Q* (null homozygote) distributions were plotted and all species were found to have 'U-shaped' distribution frequencies (*i.e.* most DArT markers showed either low or high *Q* frequencies; Figure 4.3). Accordingly, there was minimal variation between species genetic diversity estimates (mean of 2207 markers) calculated using either H_{HW} or H_{PH} (Table 4.4). Significant differences were detected between species genetic diversity estimates (Kruskal-Wallis χ^2 value 385.1, $df = 4$, $P < 0.0001$). Genetic diversity was lowest in *E. nitens* (mean H_{HW} 0.146), intermediate in *E. globulus* (mean H_{HW} 0.175) and *E. grandis* (mean H_{HW} 0.0.194) and highest in *E. camaldulensis* (mean H_{HW} 0.223) and *E. urophylla* (mean H_{HW} 0.231; Table 4.4). The same rank-order of species diversity was obtained with both H_{PH} and H_{HW} (Table 4.4).

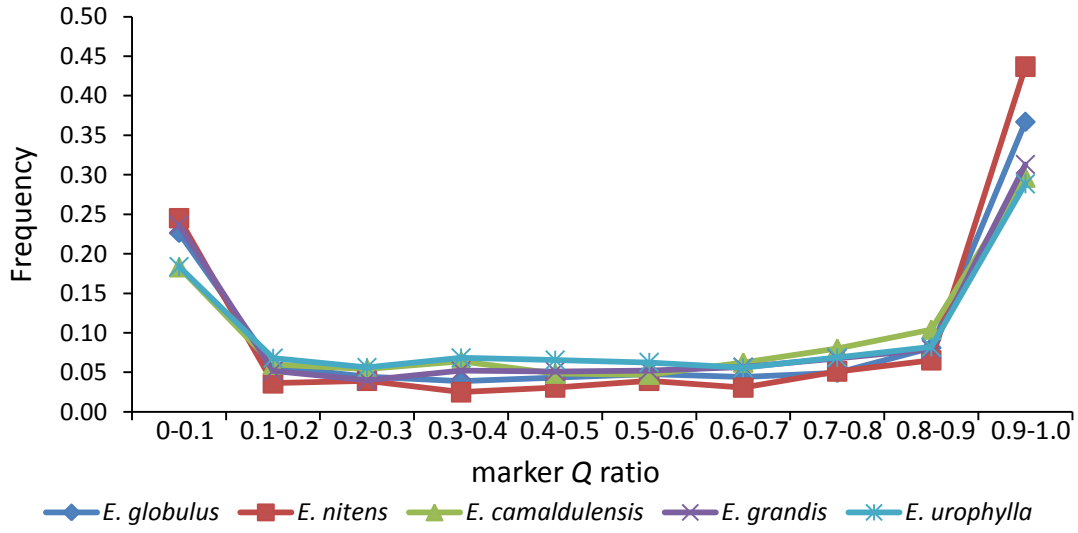

Figure 4.3 Frequency profiles of the null homozygote (*Q*) in 2207 DArT markers for five *Eucalyptus* species. All species exhibited a 'U-shaped' distribution.

Genetic diversity estimates were also calculated for each LG. Significant differences were detected between LGs using pooled data across species (Kruskal-Wallis χ^2 value 113.3, $df = 10$, $P < 0.0001$). Averaged across species, genetic diversity was lowest on LGs 11, 6 and 8, whereas the greatest diversity was detected on LGs 7, 5 and 3 (Table 4.4).

Table 4.4 Genetic diversity (*H*) for each of five *Eucalyptus* species calculated using 2207 DArT markers. Mean H_{HW} (assuming Hardy-Weinberg equilibrium) and H_{PH} (assuming no heterozygote or complete inbreeding) estimates are shown for each linkage group and for unmapped markers (number of markers shown in parentheses).

		Linkage group												
Species	Measure	1 (103)	2 (156)	3 (163)	4 (81)	5 (138)	6 (154)	7 (110)	8 (185)	9 (91)	10 (106)	11 (121)	Other[a] (799)	Average (2207)
E. nitens	H_{HW}	0.137	0.187	0.168	0.156	0.171	0.124	0.179	0.172	0.169	0.152	0.101	0.124	0.146
	H_{PH}	0.118	0.195	0.181	0.139	0.185	0.123	0.175	0.161	0.136	0.141	0.090	0.116	0.140
E. globulus	H_{HW}	0.207	0.215	0.234	0.222	0.244	0.170	0.248	0.189	0.242	0.208	0.156	0.108	0.175
	H_{PH}	0.177	0.231	0.259	0.225	0.267	0.163	0.258	0.196	0.220	0.192	0.146	0.092	0.171
E. grandis	H_{HW}	0.235	0.235	0.258	0.219	0.286	0.224	0.268	0.197	0.213	0.209	0.170	0.127	0.194
	H_{PH}	0.221	0.239	0.272	0.211	0.300	0.218	0.258	0.196	0.208	0.195	0.171	0.127	0.194
E. camaldulensis	H_{HW}	0.231	0.265	0.268	0.284	0.259	0.212	0.268	0.204	0.231	0.225	0.211	0.189	0.223
	H_{PH}	0.228	0.258	0.283	0.270	0.283	0.206	0.282	0.210	0.215	0.203	0.207	0.186	0.222
E. urophylla	H_{HW}	0.248	0.285	0.279	0.295	0.277	0.263	0.287	0.247	0.257	0.257	0.246	0.162	0.231
	H_{PH}	0.243	0.273	0.290	0.275	0.279	0.261	0.272	0.270	0.268	0.248	0.236	0.149	0.226
Average	H_{HW}	0.211	0.237	0.241	0.235	0.247	0.198	0.250	0.202	0.223	0.210	0.177	0.142	0.194
Average	H_{PH}	0.197	0.239	0.257	0.224	0.263	0.194	0.249	0.207	0.210	0.196	0.170	0.134	0.191

[a]Other; unmapped markers and the 40 multicopy markers mapped to two or more linkage groups in the *Eucalyptus* multi-species composite map.

Species genetic diversity was examined at a genome-scale by plotting H_{WH} (11 marker running-average) on the 11 LGs of the composite map (Figure 4.4 A-K). This enabled species genetic diversity to be compared at both intra and inter-section levels. A number of genetic diversity 'hot-spots' (where species exhibited lower or higher genetic diversity relative to other species) were observed; indicated by arrow symbols in Figure 4.4 A-K. Most of these hot-spots were regions of low genetic diversity, so-called 'genetic diversity deserts'. Diversity deserts were observed on multiple LGs in *E. nitens* (LG2 41 cM, Figure 4.4B; LG3 29 cM, Figure 4.4C; LG4 52 cM, Figure 4.4D; LG7 40 cM, Figure 4.4G; LG8 118 cM, Figure 4.4H; LG11 100, cM Figure 4.4K), likely reflecting the low mean genetic diversity detected in this species. However, genetic diversity deserts were also detected in *E. grandis* (LG4 51 cM, Figure 4.4D; LG8 50 cM, Figure 4.4H), *E. globulus* (LG6 109 cM, Figure 4.4F; LG11 100 cM, Figure 4.4K) and *E. camaldulensis* (LG11 100 cM; Figure 4.4K). No diversity deserts were detected in *E. urophylla*, although a region of elevated genetic diversity was observed in this species on LG10 (~40 cM; Figure 4.4J).

Following pages:

Figure 4.4 A-K Species genome-wide genetic diversity (H_{HW}). An 11 marker running average is plotted for each linkage group; using 1408 markers of the five species dataset mapped on the *Eucalyptus* multi-species composite linkage map. Six graphs are shown for each linkage group; one graph per species and a combined species graph. Triangle symbols above *E. globulus*, *E. grandis* and *E. camaldulensis* graphs indicate the linkage group positions of markers used to calculate the H_{HW} 11-marker running average. The final marker on each linkage group (indicating the length of the composite map linkage group) is indicated by either, (1) circle symbol (marker included in genetic diversity calculations), or (2) cross symbol (not included in calculations). Arrows highlight regions of lower or elevated genetic diversity in a species relative to either, (1) all species, or (2) species of the same section (for *Maidenaria* and *Latoangulatae* species only).

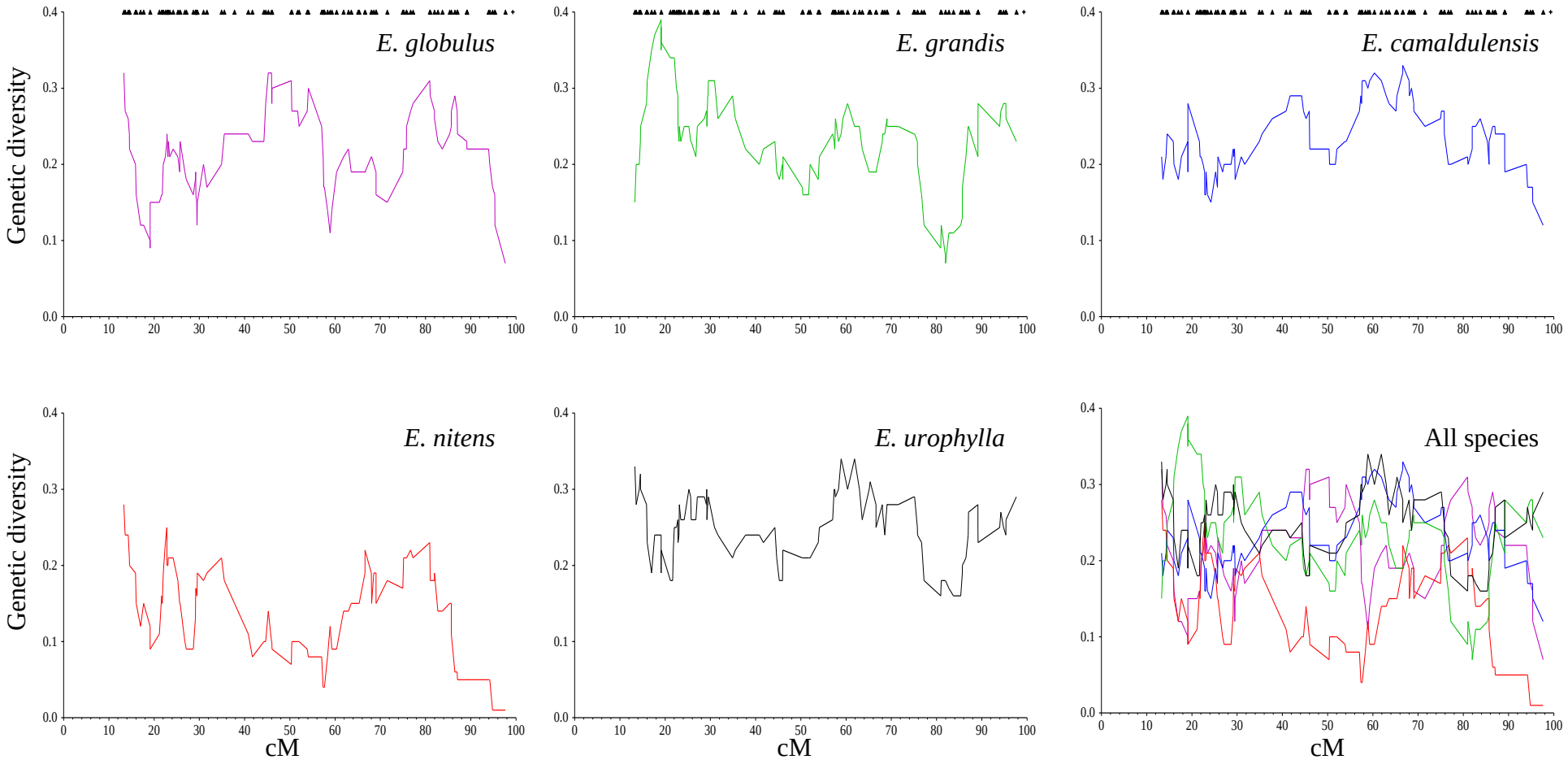

Figure 4.4A Linkage group 1 species genetic diversity (H_{HW}).

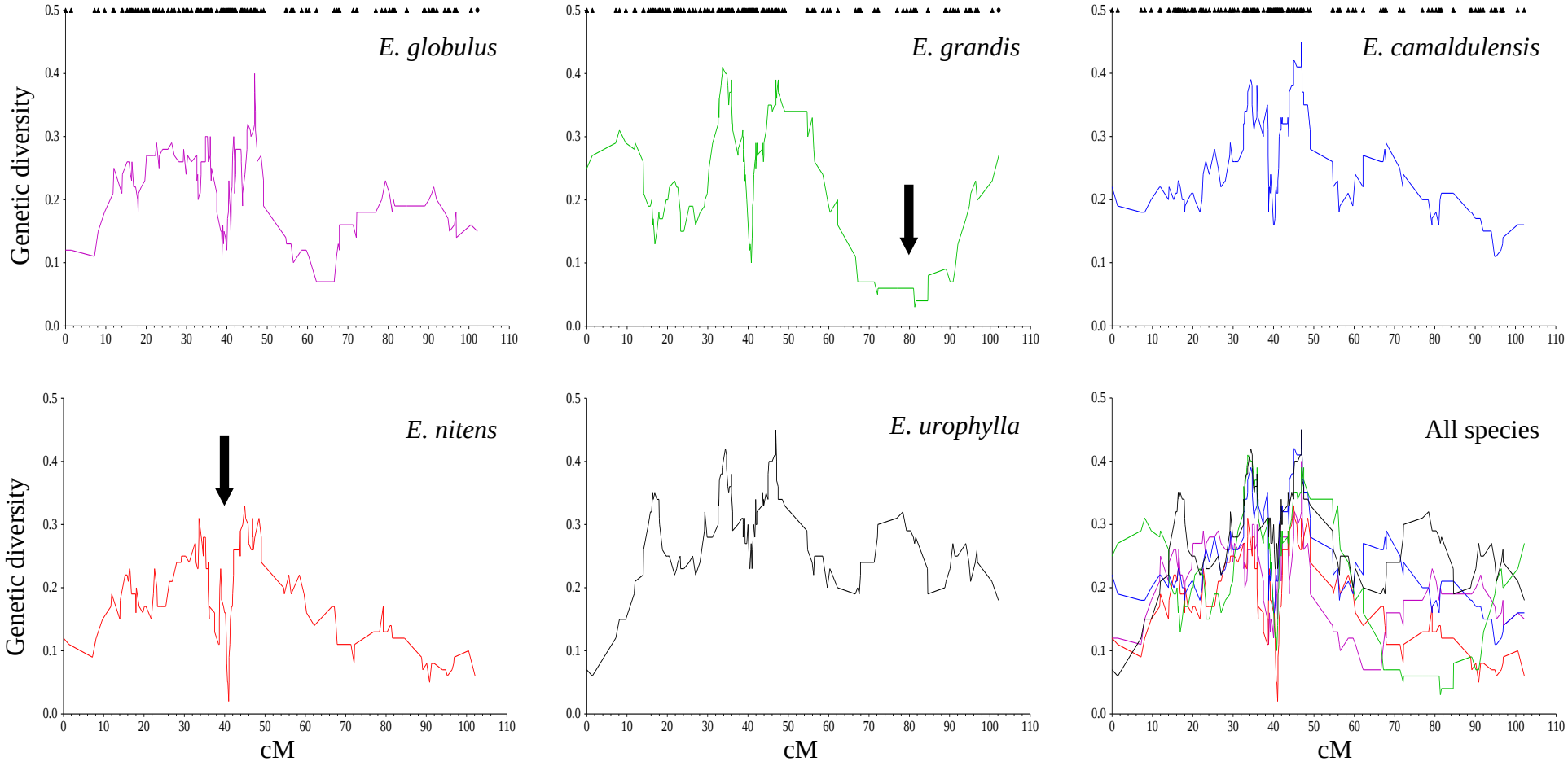

Figure 4.4B Linkage group 2 species genetic diversity (H_{HW}).

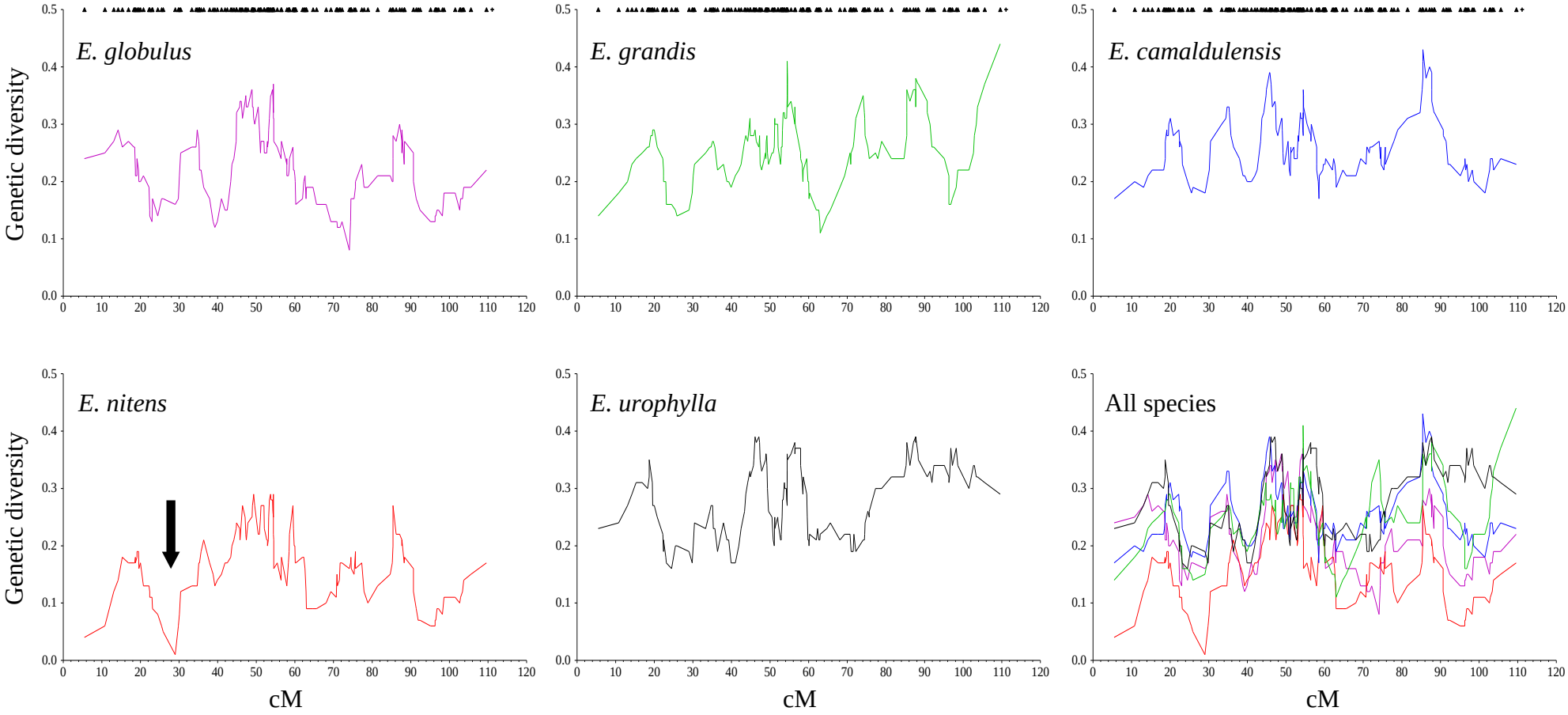

Figure 4.4C Linkage group 3 species genetic diversity (H_{HW}).

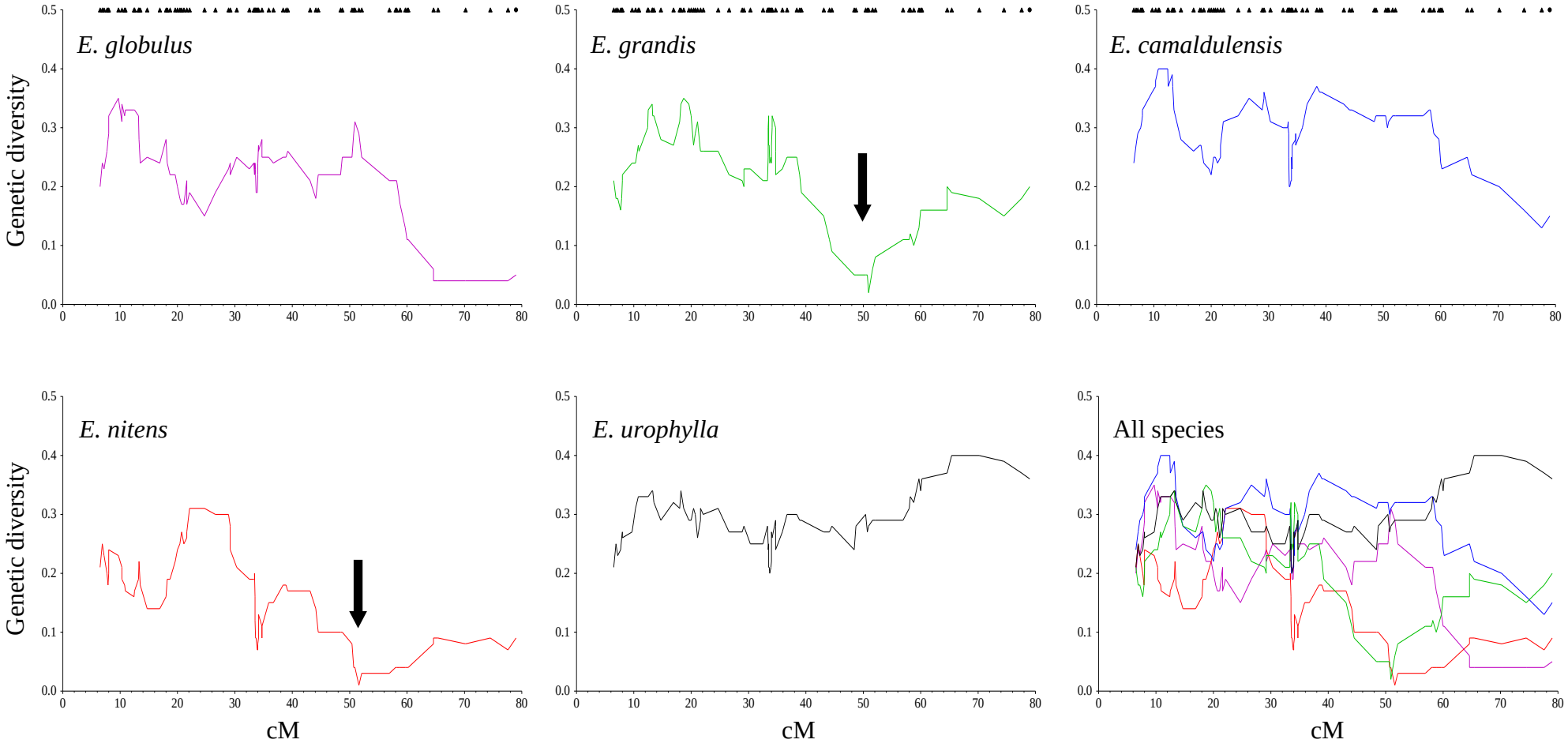

Figure 4.4D Linkage group 4 species genetic diversity (H_{HW}).

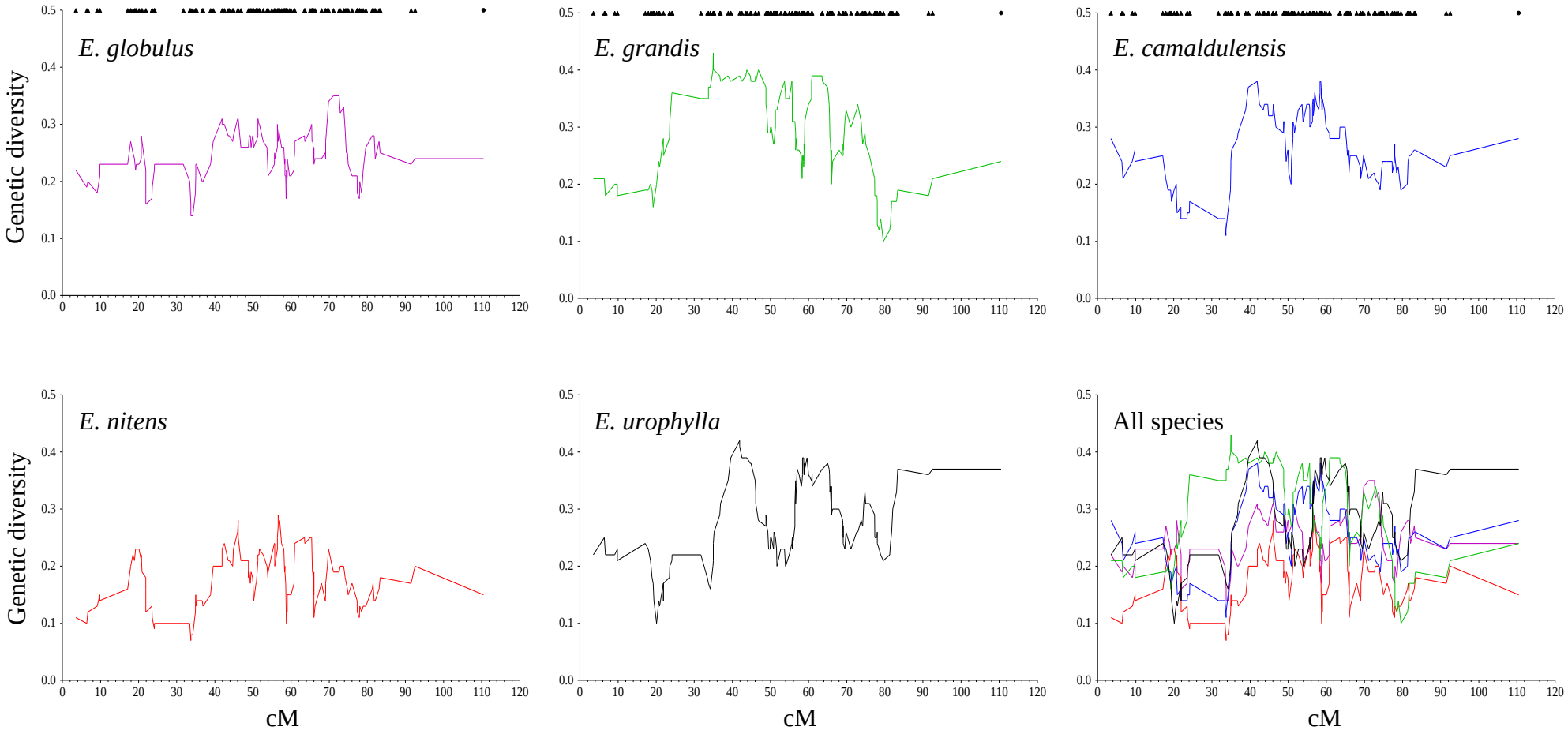

Figure 4.4E Linkage group 5 species genetic diversity (H_{HW}).

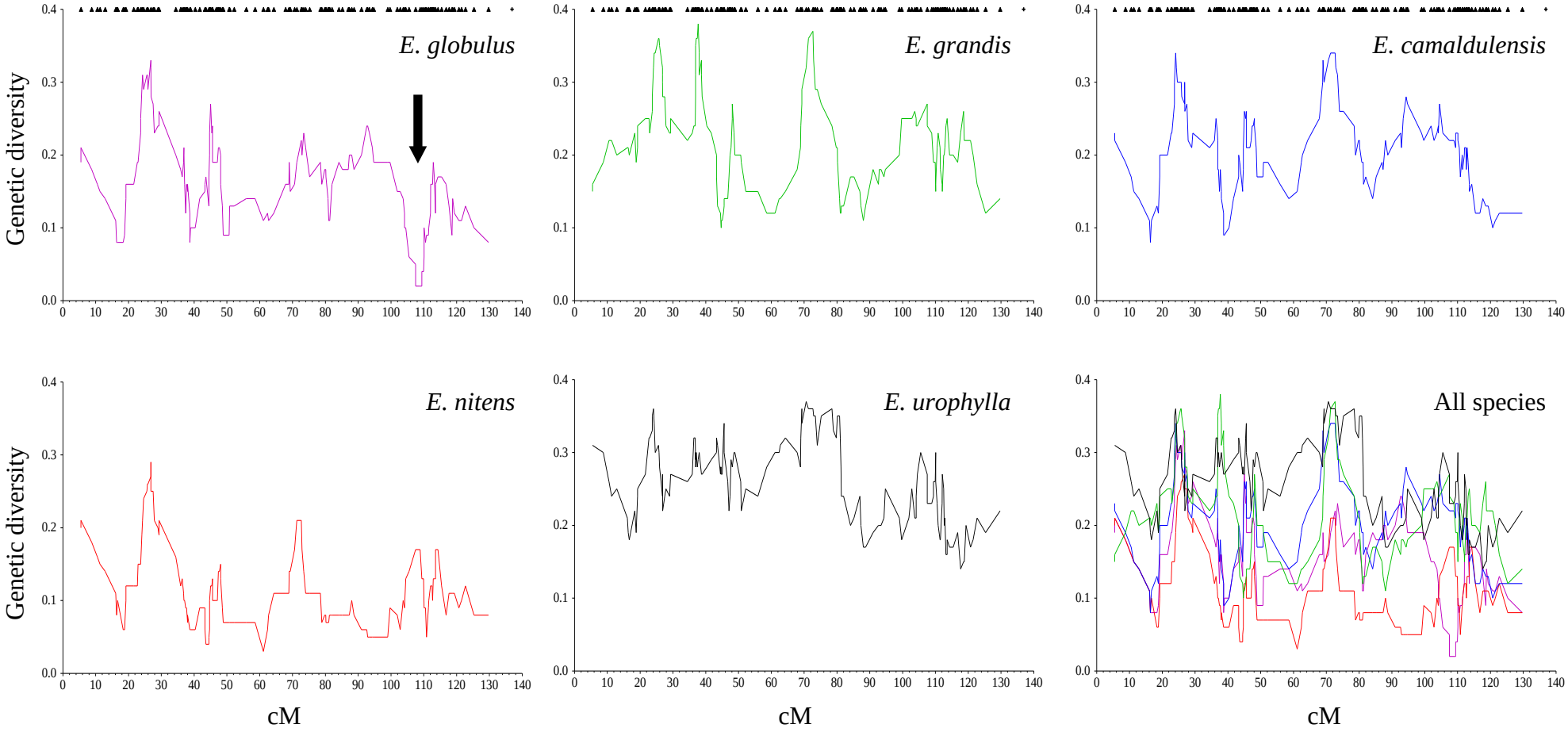

Figure 4.4F Linkage group 6 species genetic diversity (H_{HW}).

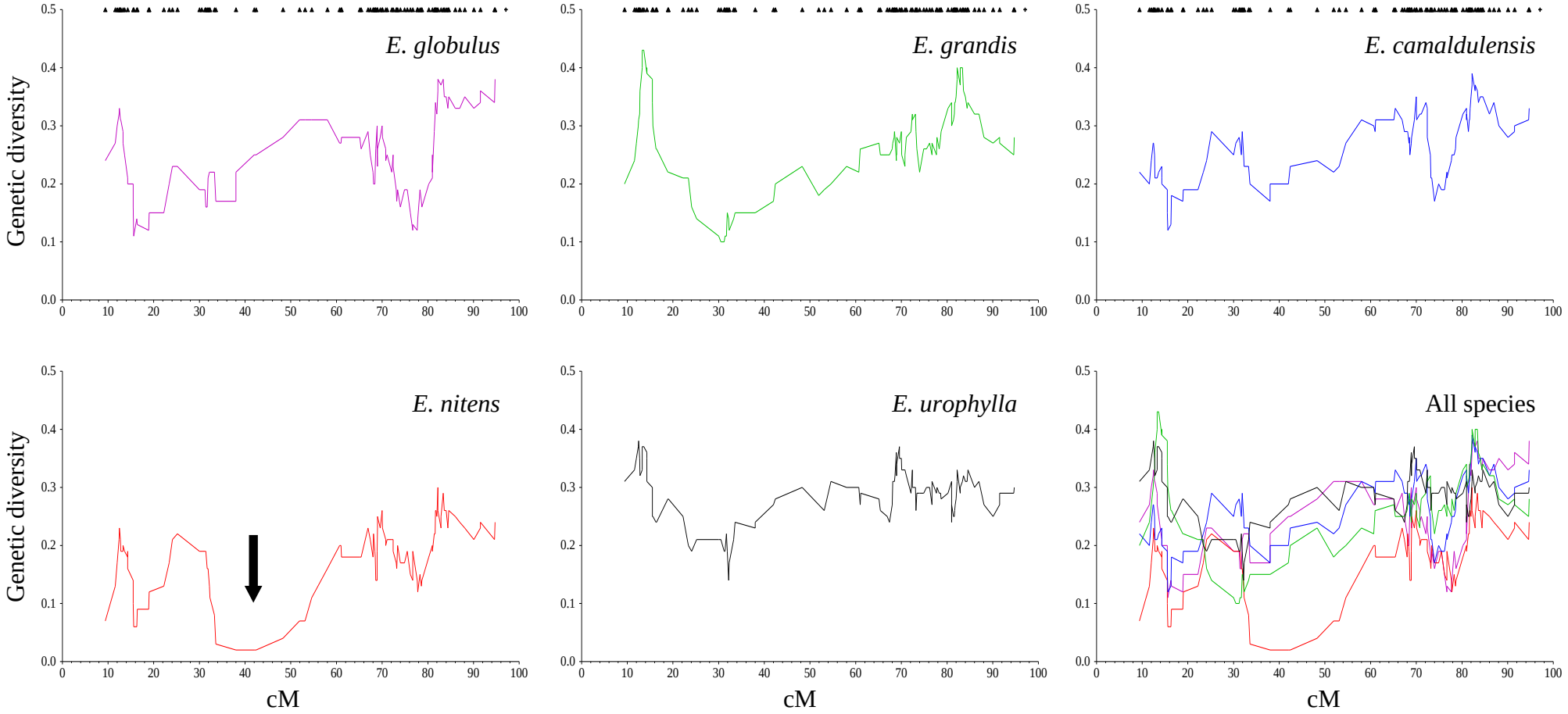

Figure 4.4G Linkage group 7 species genetic diversity (H_{HW}).

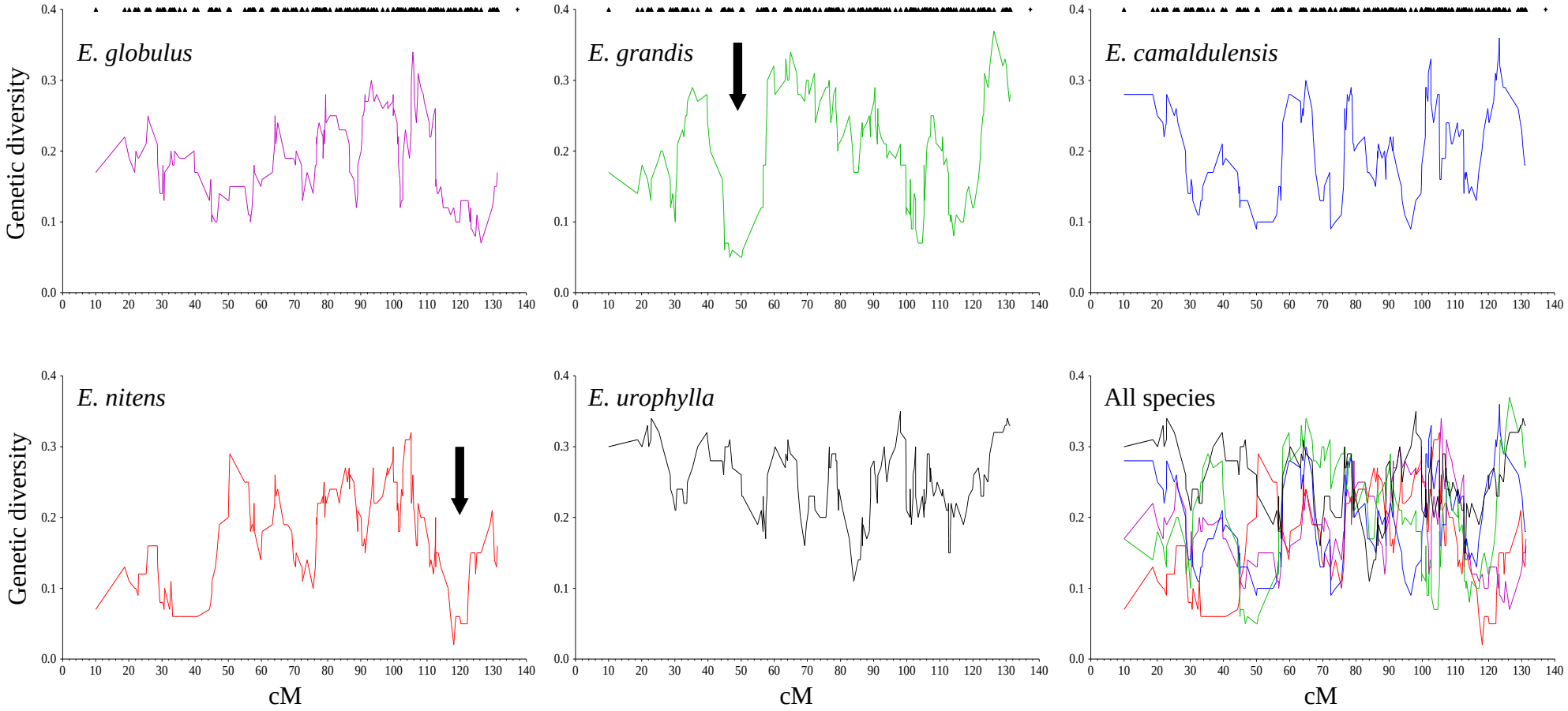

Figure 4.4H Linkage group 8 species genetic diversity (H_{HW}).

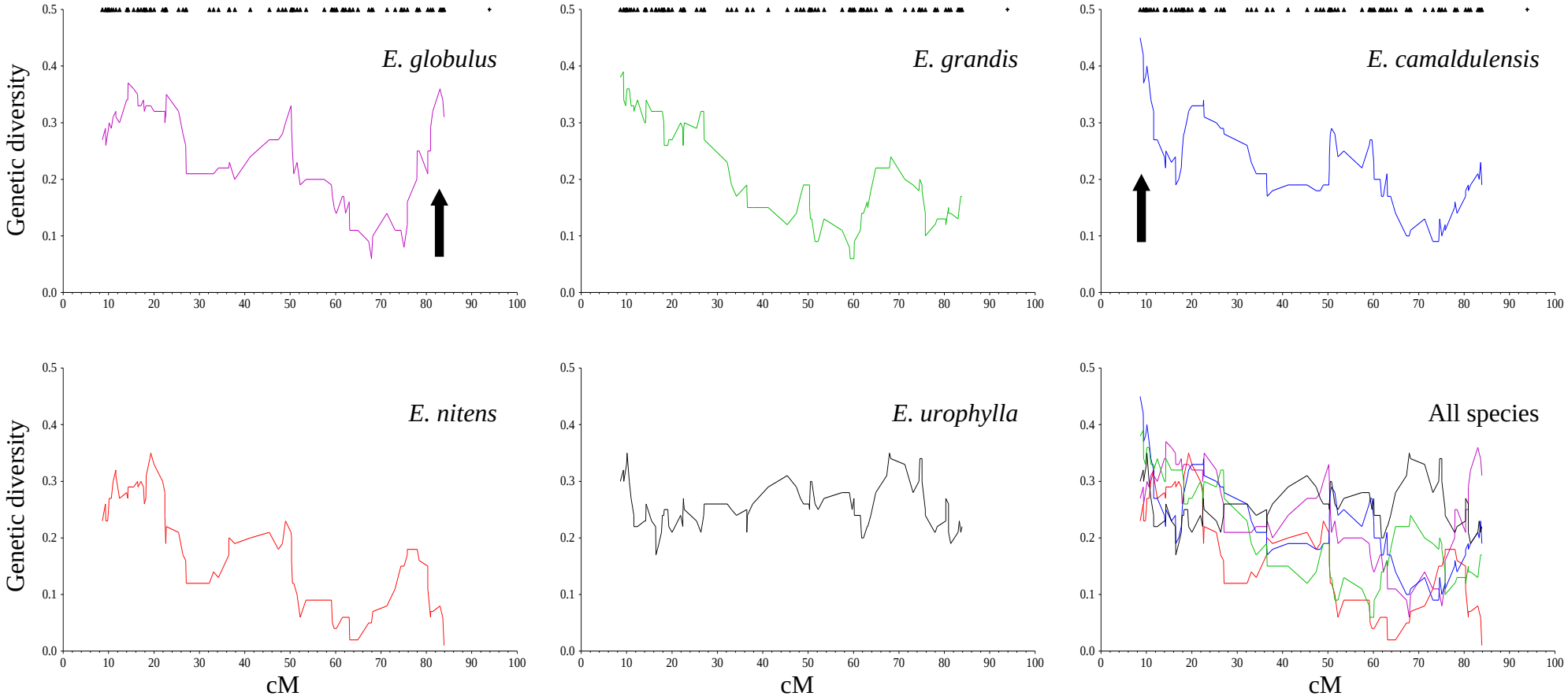

Figure 4.4I Linkage group 9 species genetic diversity (H_{HW}).

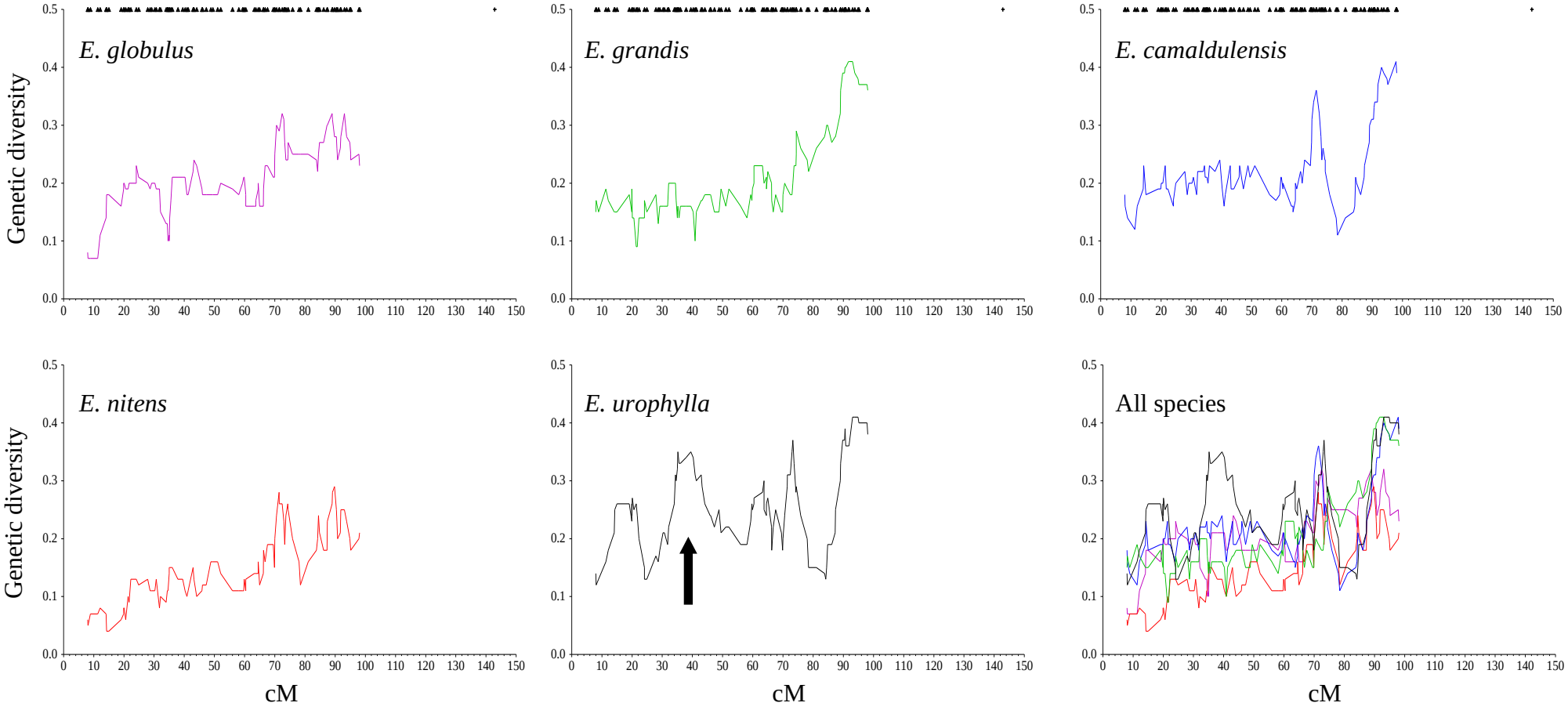

Figure 4.4J Linkage group 10 species genetic diversity (H_{HW}).

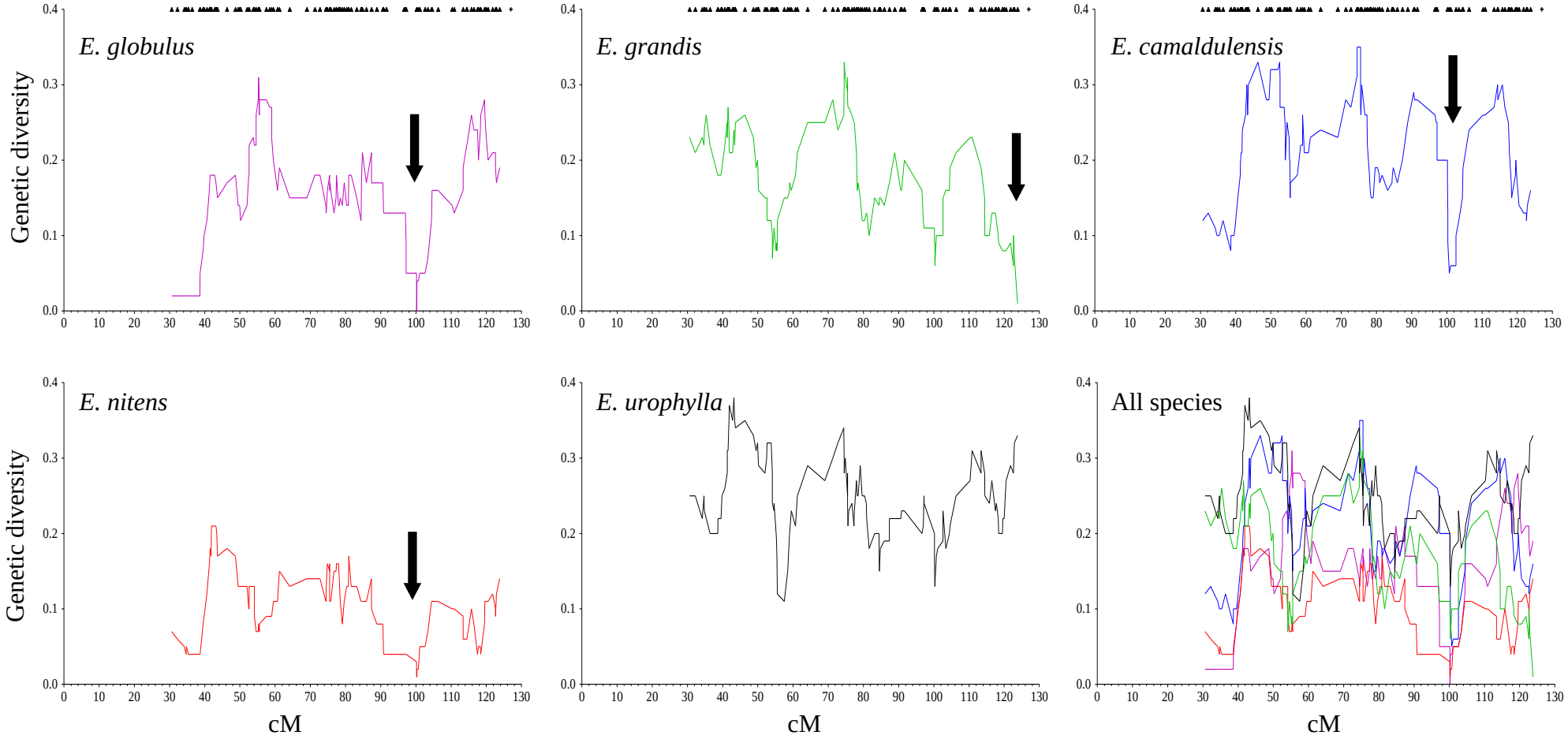

Figure 4.4K Linkage group 11 species genetic diversity (H_{HW}).

4.3.5 Fixed marker differences and Bayescan outlier analyses

Fixed marker differences (1 ratio = 1; all individuals scored as '1' in one species and as '0' in a second species) were detected in all 10 pair-wise species comparisons and ranged from 1 (*E. grandis* vs. *E. urophylla*) to 62 fixed markers (*E. nitens* vs. *E. grandis* comparison; Table 4.5). A greater number of fixed marker differences were typically found between inter-section species comparisons (especially involving section *Maidenaria* species) in comparison to intra-section comparisons (Table 4.5). In Bayescan analyses, markers having '1 scores' in < 2% and ≥ 98% of individuals in each pair-wise species comparison were removed from the 2207 marker dataset. Thus, Bayescan analyses were conducted on a total of 1483 (*E. globulus* vs. *E. nitens*) to 1898 markers (*E. urophylla* vs. *E. camaldulensis*; Table 4.5). Mean marker F_{ST} values were high in all comparisons and ranged from 0.26 (*E. urophylla* vs. *E. camaldulensis*) to 0.44 (*E. nitens* vs. *E. grandis*; Table 4.5). The rank order of pair-wise species F_{ST} estimates calculated using Bayescan was very similar to the rank order of pair-wise species F_{ST} estimates calculated using GenAlEx (using the 2207 marker dataset). The number of fixed marker differences detected between species-pairs reflected their degree of differentiation; e.g. higher F_{ST} values equalled more fixed marker differences. The number and significance level of outlier markers detected in each comparison also appeared to be influenced by the level of species differentiation. For example, species differentiation was high between *E. nitens* and both *Latoangulatae* species (*E. grandis*, Bayescan mean F_{ST} = 0.44, and *E. urophylla* Bayescan mean F_{ST} = 0.40; Table 4.5). In both of these comparisons, only fixed markers were detected as outlier markers and all of these markers were significant at the 'barely worth mentioning' threshold only. In contrast, outlier markers were detected at higher significance thresholds (e.g. 'decisive' level) in less differentiated species comparison (e.g. *E. urophylla* vs. *E. camaldulensis* comparison, Bayescan mean marker F_{ST} = 0.26; Table 4.5). Markers having a 1 ratio difference as low as 0.90 were also detected as outliers in the *E. urophylla* vs. *E. camaldulensis* comparison (Table 4.5).

In all species comparisons, outlier marker F_{ST} estimates were greater than the pair-wise species mean, suggesting that these outlier markers were under directional selection. The number of outlier markers detected in each comparison ranged from 30 (*E. grandis* vs. *E. urophylla*; 1.7% markers analysed) to 62 (*E. nitens* vs. *E. grandis*, 3.5% markers analysed; Table 4.5) and across all comparisons totalled 447 markers. Unfortunately, not all outlier markers were mapped on the composite map, with the proportion of

mapped markers ranging from 29% (*E. nitens* vs. *E. grandis*) to 71% (*E. urophylla* vs. *E. camaldulensis*; Table 4.5) in each of the ten comparisons.

Table 4.5 Number of fixed marker differences (1 ratio = 1), diagnostic markers (1 ratio ≥ 0.90) and outlier markers detected in Bayescan analyses in each pair-wise species comparison. In Bayescan analyses, *n* equals the number of markers following the removal of markers with 1 scores in < 2% and ≥ 98% of individuals. Mean marker F_{ST}, significance level of outlier markers (Jeffrey's scale; see materials and methods) and the number and percentage of outlier markers mapped on the *Eucalyptus* multi-species composite map are shown.

Comparison		1 ratio difference (threshold and number of markers)							Markers detected as 'outliers' in Bayescan analyses[a]								
Section[b]	Species[c]	1	≥ 0.99	≥ 0.98	≥ 0.97	≥ 0.96	≥ 0.95	≥ 0.90	*n*	Mean F_{ST}	Dec	Very strong	Strong	Sub	BWM	Total (%)[d]	Mapped Mks (%)[e]
Maid vs. *Lat*	GL vs. GR	41	41	69	90	108	133	173	1790	0.40				41	5	46 (2.6)	17 (37.0)
	GL vs. U	25	25	45	59	66	87	126	1895	0.38				25	16	41 (2.2)	27 (65.9)
	N vs. GR	62	62	92	110	134	149	211	1766	0.44					62	62 (3.5)	18 (29.0)
	N vs. U	34	34	55	67	80	100	165	1862	0.40					34	34 (1.8)	27 (79.4)
Maid / Lat vs. *Ex*	GL vs. C	38	38	57	67	87	105	167	1870	0.37			6	32	19	57 (3.0)	24 (42.1)
	N vs. C	48	48	74	96	113	130	202	1878	0.40				4	44	48 (2.6)	28 (58.3)
	GR vs. C	12	12	26	34	49	56	106	1883	0.32		11	3	15	21	50 (2.7)	15 (30.0)
	U vs. C	2	2	6	13	15	20	42	1898	0.26	4	3	7	10	17	41 (2.2)	29 (70.7)
Intra-section	GR vs. U	1	1	4	12	13	22	46	1804	0.28	1	3	3	11	12	30 (1.7)	13 (43.3)
	GL vs. N	15	15	25	34	40	43	69	1493	0.33		11	4	11	12	38 (2.6)	17 (44.7)
	Total	278	278	453	582	705	845	1307			5	28	23	149	242	447	215 (48.1)

[a]Bayescan significance thresholds; Dec = decisive, Sub = substantial and BWM = barely worth mentioning (posterior probabilities given in materials and methods). [b]Section; *Maid* = *Maidenaria*, *Lat* = *Latoangulatae*, *Ex* = *Exsertaria*. [c]Species; GL = *E. globulus*, GR = *E. grandis*, U = *E. urophylla*, N = *E. nitens* and C = *E. camaldulensis*. [d]Total; percentage of markers detected as outliers, calculated using *n*. [e]Mapped Mks (markers); number and percentage of 'outlier' markers mapped on the *Eucalyptus* multi-species composite linkage map.

Many markers were detected as outliers in two or more comparisons (see Figure 4.5A-K). The 447 outlier markers detected in all comparisons comprised 228 unique markers (Table 4.6). Of these unique markers, approximately half (119) were mapped on the composite map; these represented 215 outlier markers across all comparisons (Table 4.6 and Figure 4.5A-K). The number of unique outlier markers occurring on each LG ranged from 2 (LG3) to 22 (LG11). Of the 1408 markers on the composite map, the number of outlier markers detected represented 1.2% to 18.2% of the markers on each LG (Table 4.6).

Table 4.6 Number of outlier markers detected on each linkage group in Bayescan analyses. The total number of mapped (by linkage group) and unmapped outlier markers detected across all ten pair-wise species comparisons are shown. Markers were often detected as outliers in multiple species comparisons; 'Unique markers' indicate the number of unique markers detected across all 10 comparisons. The percentage of markers detected as outliers are shown for both mapped markers (1408; by linkage group) and unmapped markers (799).

Linkage group	Outlier markers	Unique markers	2207 marker dataset markers and percentage of unique outlier markers
1	8	6	103 (5.8)
2	33	17	156 (10.9)
3	6	2	163 (1.2)
4	24	13	81 (16.0)
5	9	5	138 (3.6)
6	20	13	154 (8.4)
7	5	3	110 (2.7)
8	37	18	185 (9.7)
9	7	7	91 (7.7)
10	24	13	106 (12.3)
11	42	22	121 (18.2)
Subtotal	215	119	1408 (8.5)
Unmapped	232	109	799 (13.6)
Total	447	228	2207 (10.3)

A number of outlier marker 'clusters', in which outlier markers (≥ 3 unique outlier markers across all comparisons) resided within small chromosomal regions (≤ 5 cM), were found within LGs. Single clusters were detected on LGs 2, 4, 5 and 10 (marked by horizontal bars in Figure 4.5 B/D/E/J). The high number of unique outlier markers

detected on LG8 (18 unique outlier markers; Table 4.6 and Figure 4.5H) and LG11 (22 unique outlier markers; Table 4.6 and Figure 4.5K) were found to group in multiple clusters on these LGs. In the outlier cluster on LG11 (96 to 103 cM; Figure 4.5K), significant outlier markers were detected in all pair-wise species comparisons. Within this region, three to five outlier markers were detected in each of four pair-wise species comparisons (marked with arrow symbols in Figure 4.5K); including the *Latoangulatae* intra-section *E. grandis* vs. *E. urophylla* comparison in which five outlier markers were detected. All species exhibited a reduction in genetic diversity in this LG11 chromosomal region, with pronounced reductions occurring in *E. globulus*, *E. nitens* and *E. camaldulensis* (see Figure 4.4K), suggesting that this region may be a hot-spot of species differentiation. The outlier cluster identified on LG2 (Figure 4.5B) similarly coincided with a region of low genetic diversity in *E. nitens* and *E. grandis* (Figure 4.4B). The remaining outlier clusters did not co-locate with genetic diversity deserts.

Following pages:

Figure 4.5 A-K Linkage group distribution of outlier markers detected in all 10 pair-wise species comparisons. Diamond symbols indicate the positions of outlier markers with their significance indicated by asterisks' above each marker; * = barely worth mentioning, ** = substantial, *** = strong, **** = very strong and ***** = decisive. Where outlier markers overlap, significance indicators are stacked vertically above marker symbols. Markers identified as outlier markers in more than one comparison are indicated by letters (e.g. markers labelled with an 'a' are identical; marker names given in the title of each figure). Linkage group regions containing outlier marker clusters (three or more unique outlier markers across all comparisons) are highlighted with horizontal bars. Arrows indicate outlier clusters within species comparisons. Linkage group length (cM) is plotted on the horizontal axis, with each horizontal dashed line representing a pair-wise species comparison. Comparisons are group by species within inter or intra-section groupings; *Lat* = section *Latoangulatae*, *Maid* = section *Maidenaria*, *Ex* = section *Exsertaria*, GL = *E. globulus*, GR = *E. grandis*, U = *E. urophylla*, C, = *E. camaldulensis* and N = *E. nitens.*

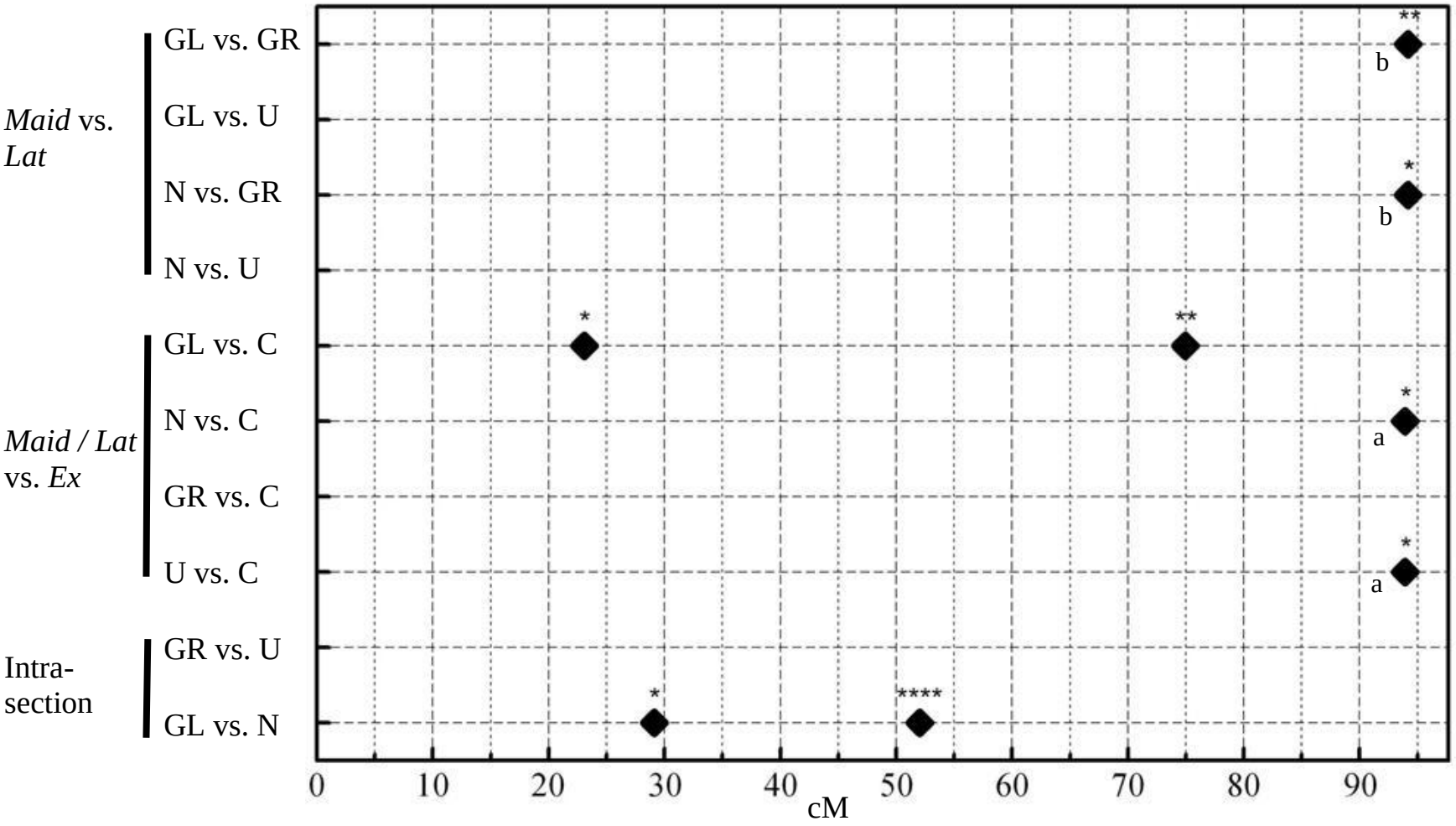

Figure 4.5A Linkage group 1; a) ePt-600709, b) ePt-641800.

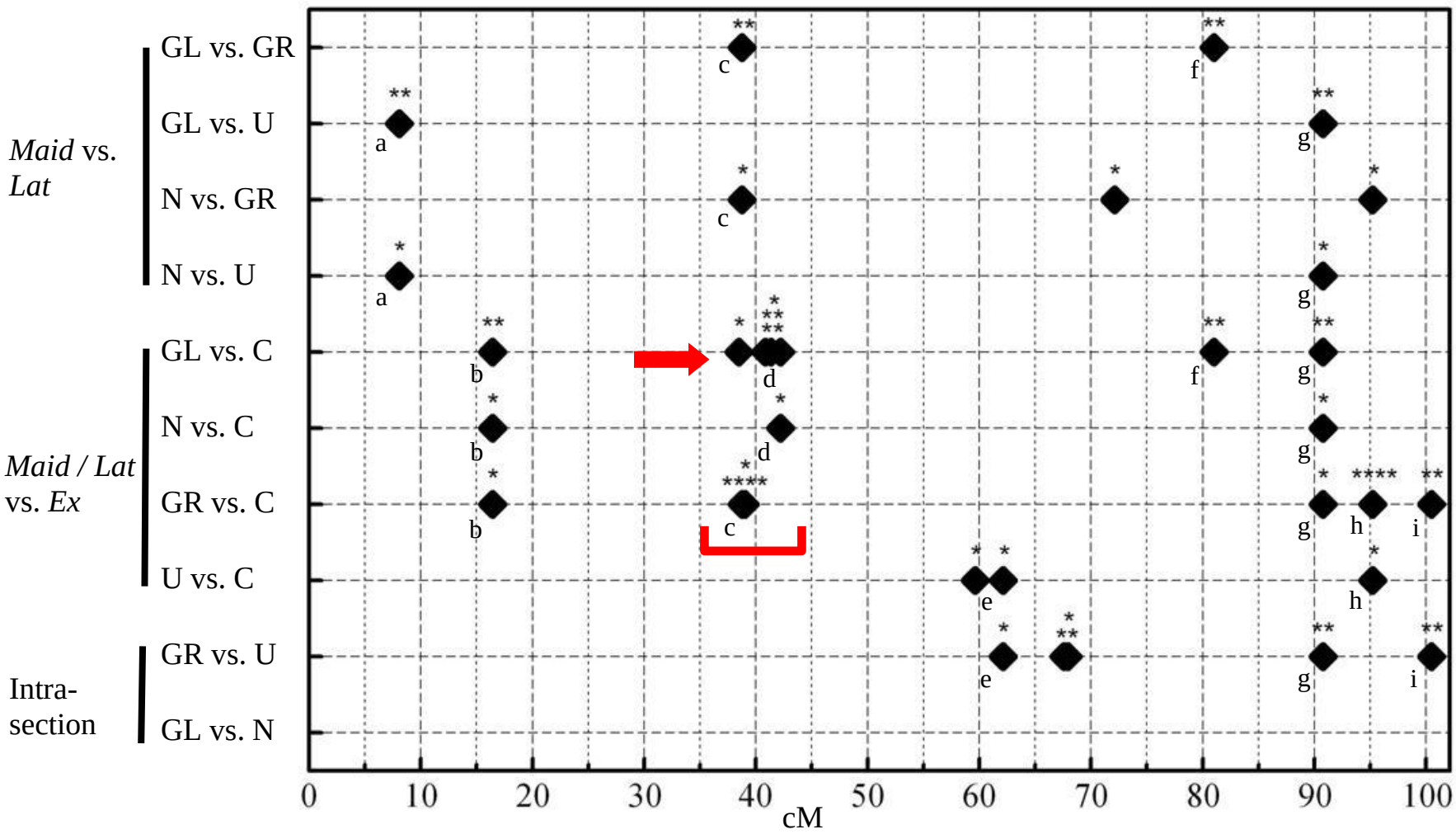

Figure 4.5B Linkage group 2; a) ePt-643799, b) ePt-634991, c) ePt-574488, d) ePt-599897, e) ePt-641504, f) ePt-643956, g) ePt-644128, h) ePt-640029 and i) ePt-599530.

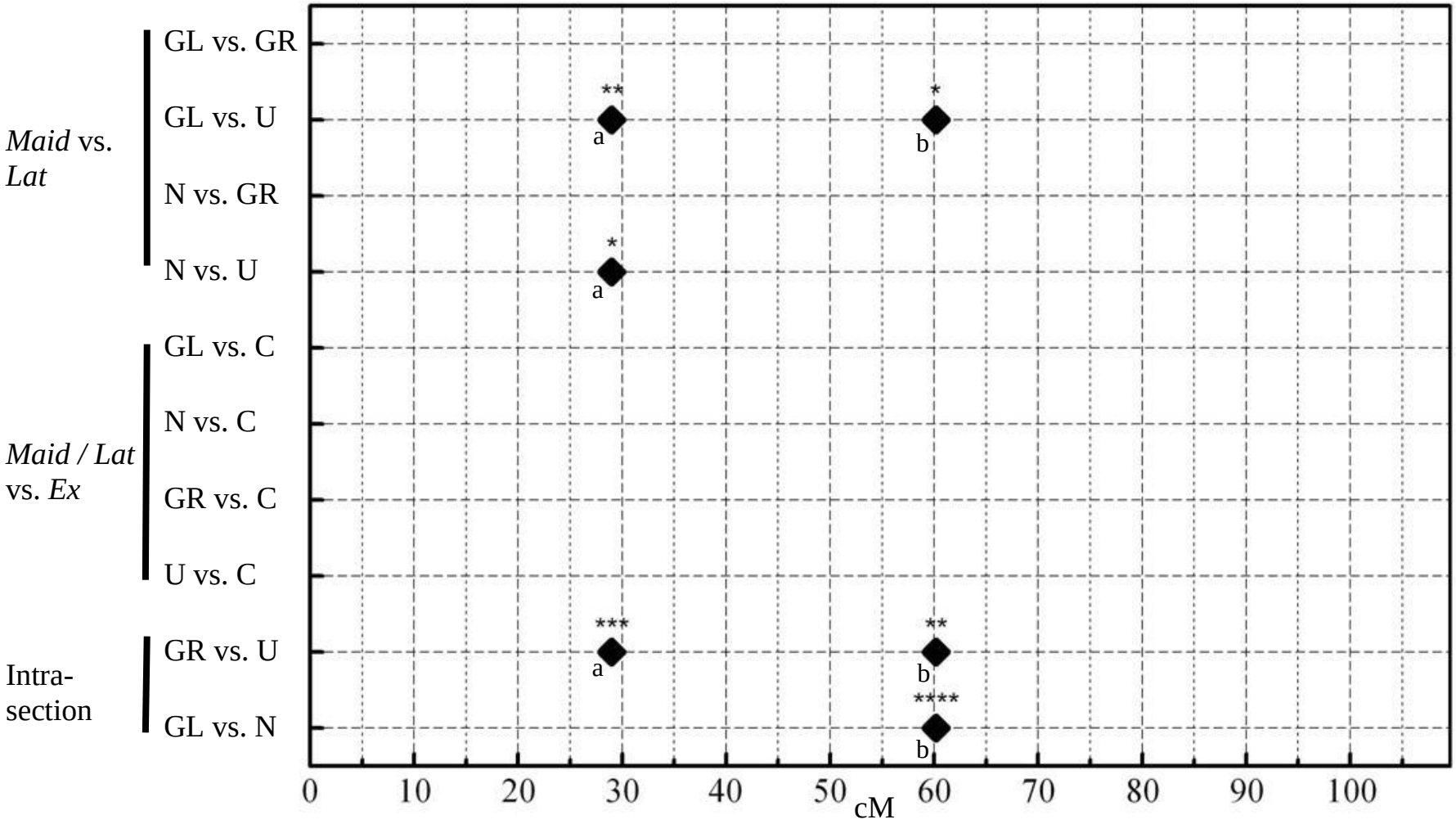

Figure 4.5C Linkage group 3; a) ePt-575216 and b) ePt-637106.

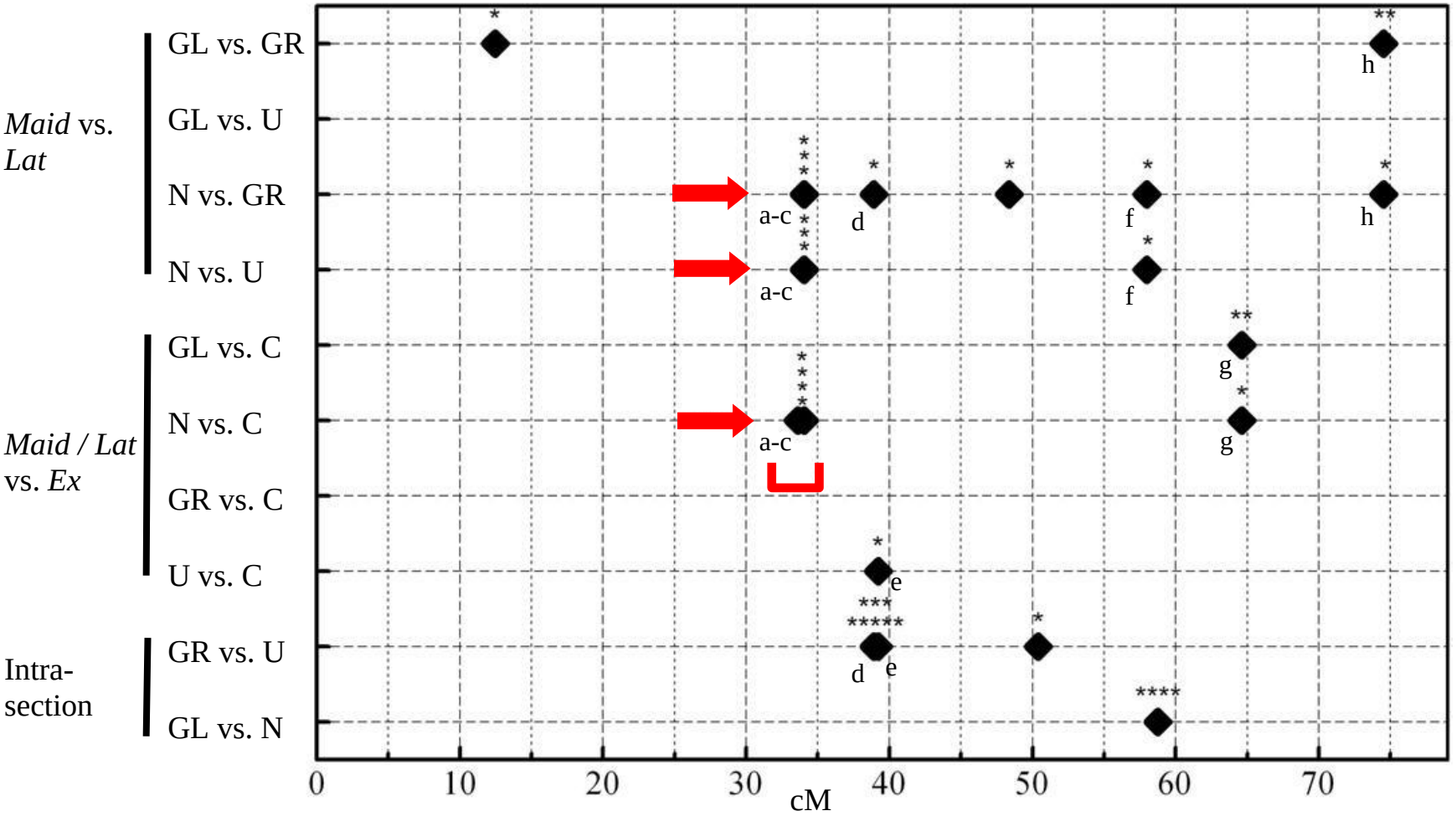

Figure 4.5D Linkage group 4; a-c) ePt-571640, ePt-564008 and ePt-641176, d) ePt-638305, e) ePt-644182, f) ePt-570682, g) ePt-574558 and h) ePt-504250.

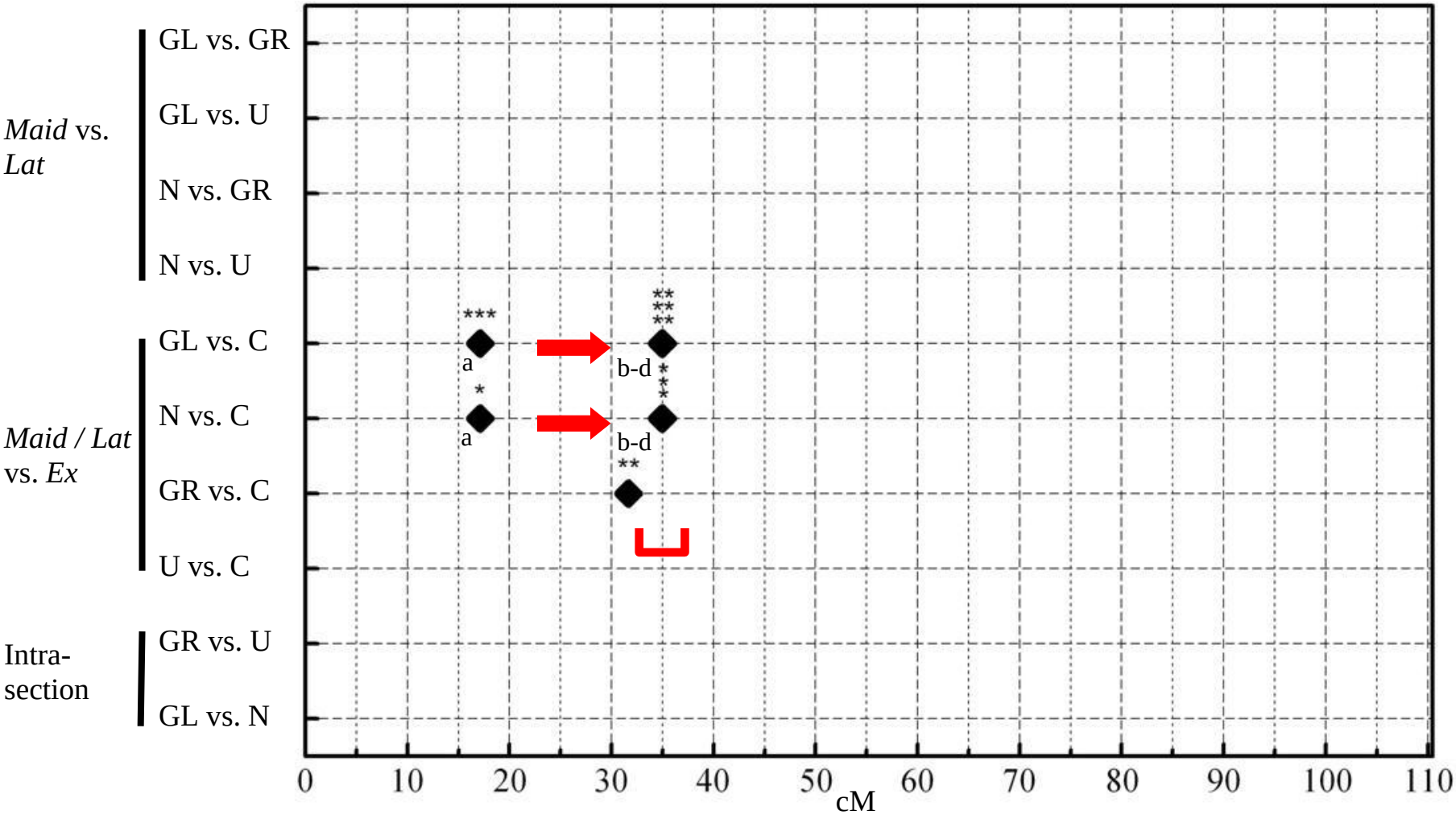

Figure 4.5E Linkage group 5; a) ePt-573579 and b-d) ePt-639853, ePt-600341 and ePt-570274.

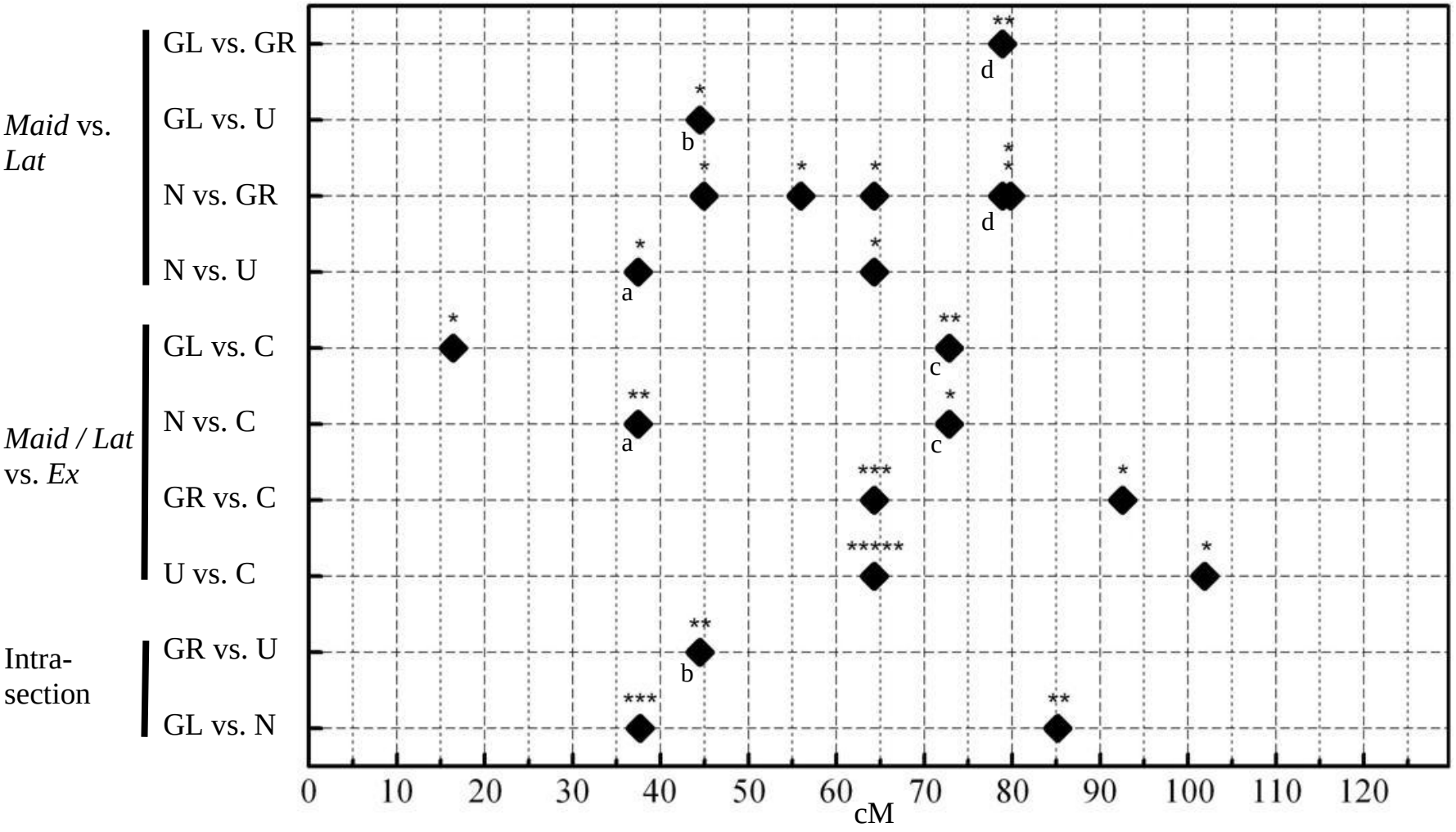

Figure 4.5F Linkage group 6; a) ePt-643644, b) ePt-642976, c) ePt-571726 and d) ePt-572381.

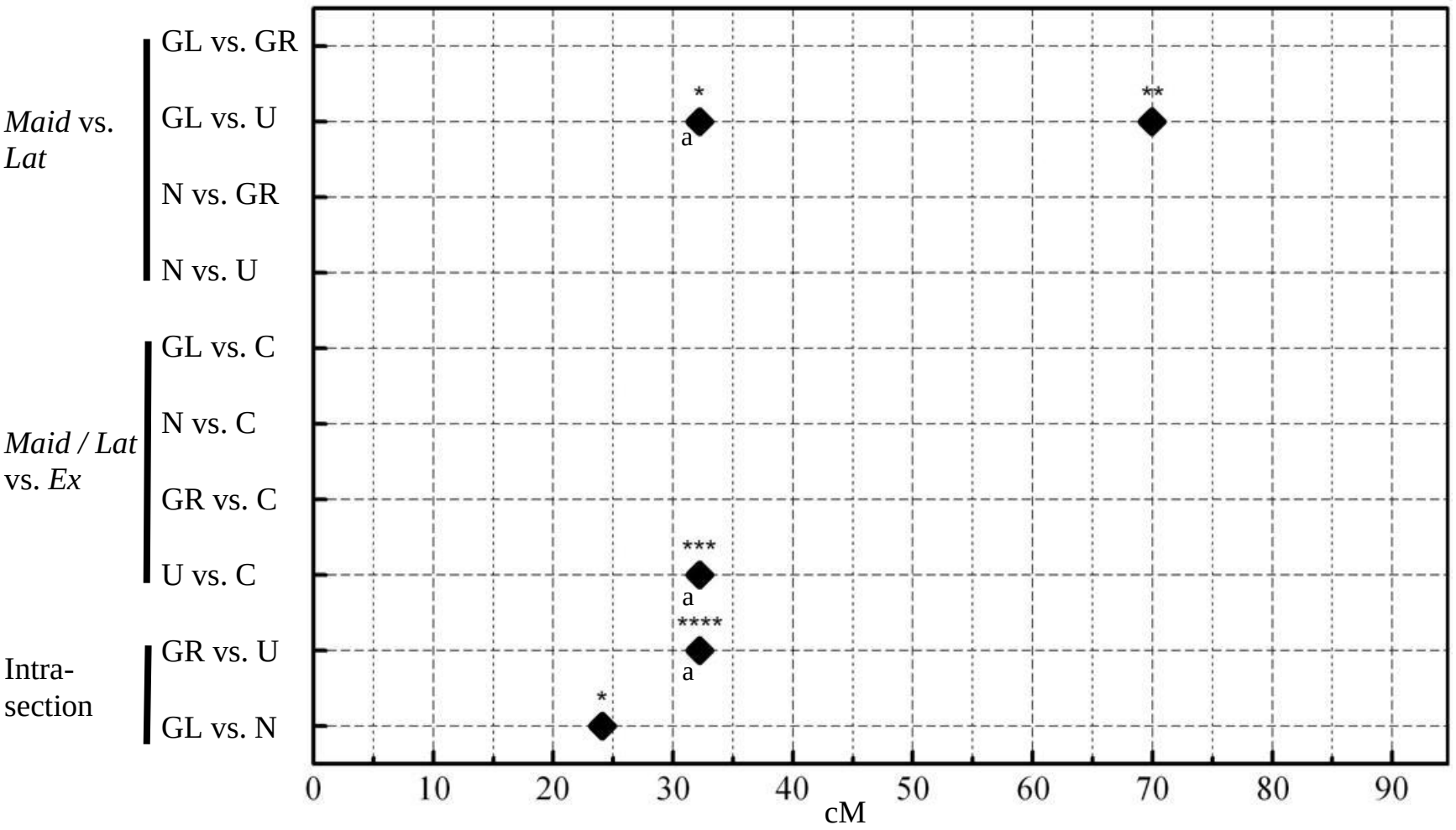

Figure 4.5G Linkage group 7; a) ePt-643010.

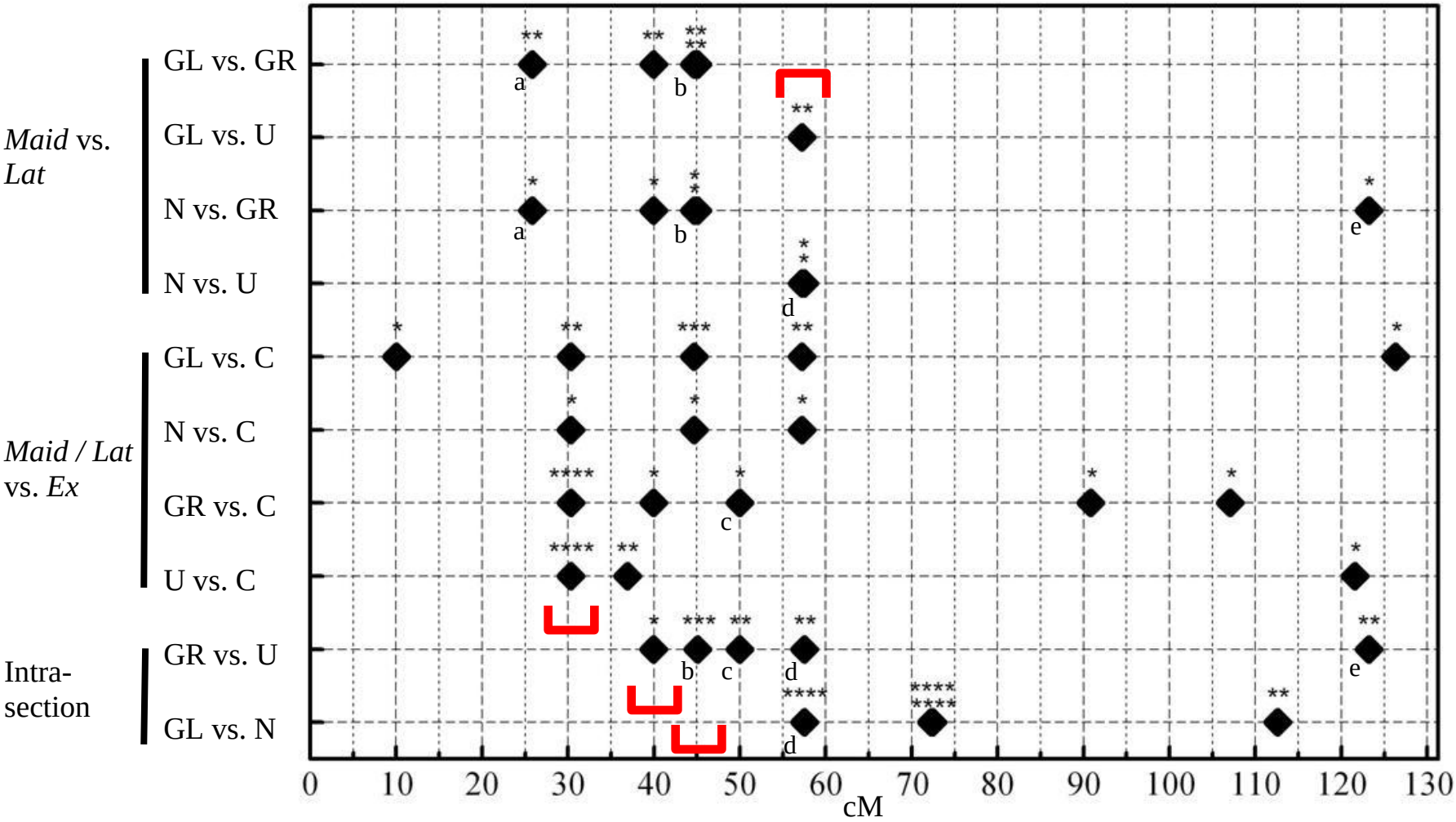

Figure 4.5H Linkage group 8; a) ePt-574967, b) ePt-642994, c) ePt-642970, d) ePt-642862 and e) ePt-571631.

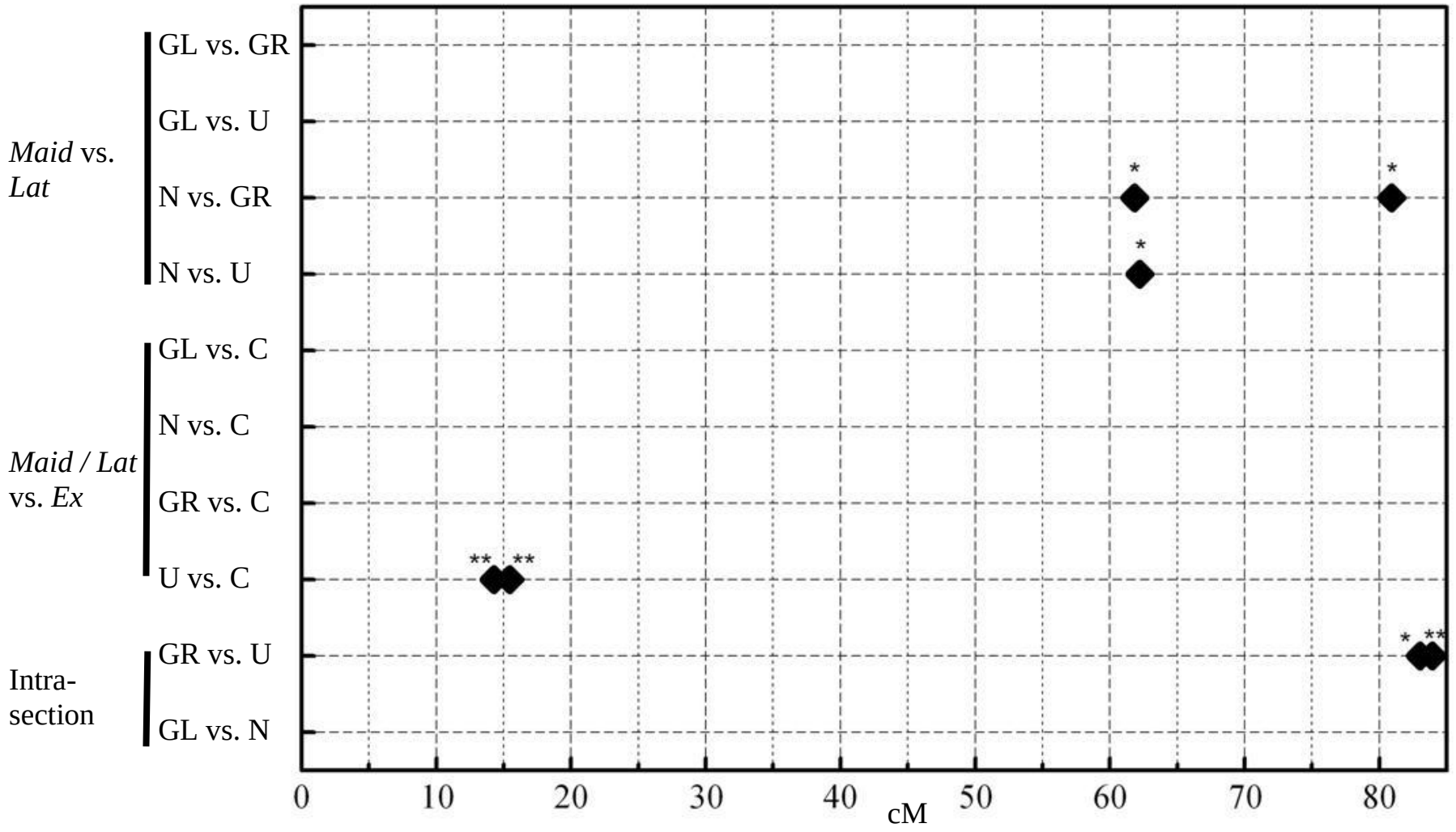

Figure 4.5I Linkage group 9.

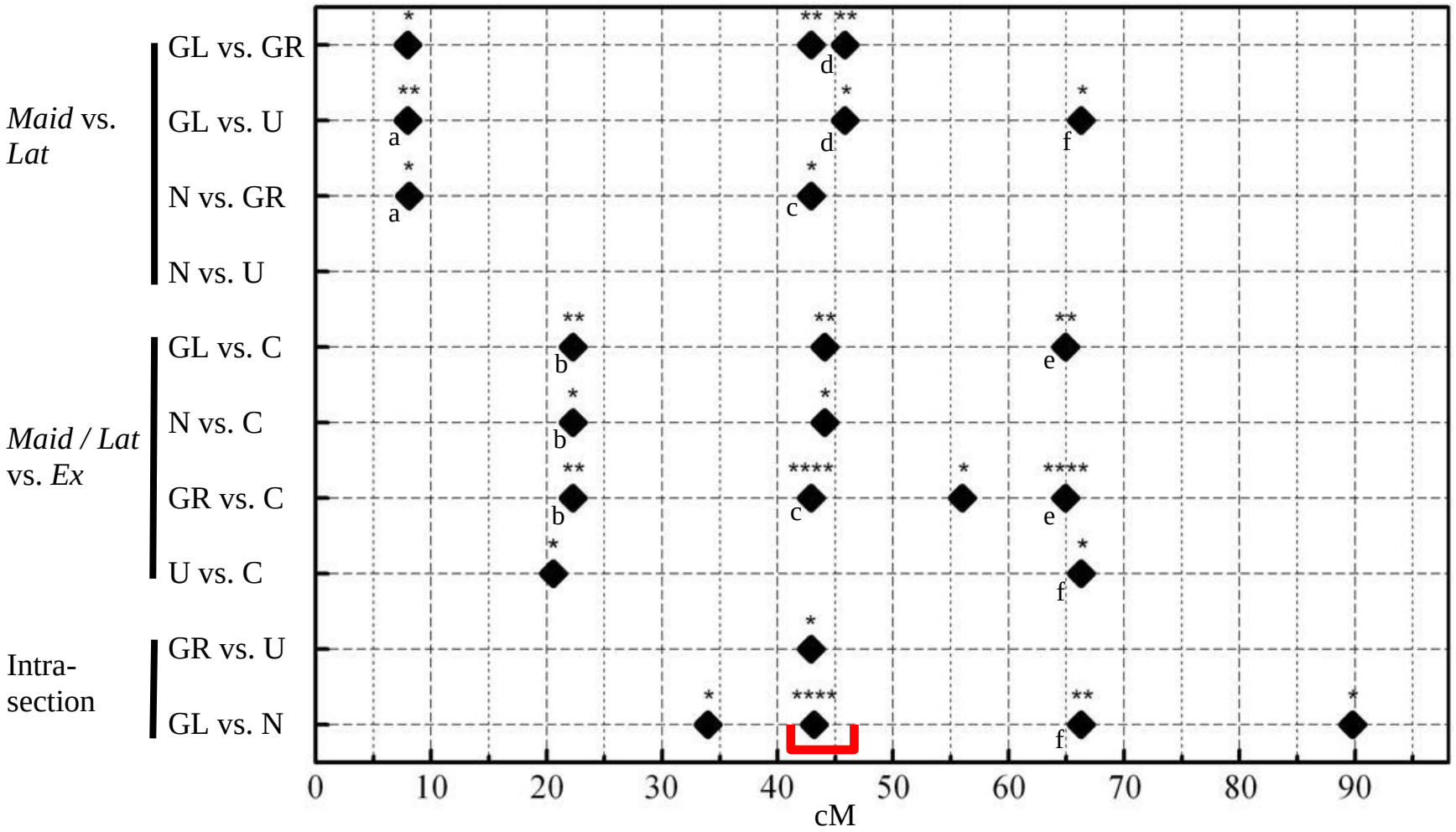

Figure 4.5J Linkage group 10; a) ePt-639246, b) ePt-639336, c) ePt-568620, d) ePt-600541, e) ePt-570076 and f) ePt-574739.

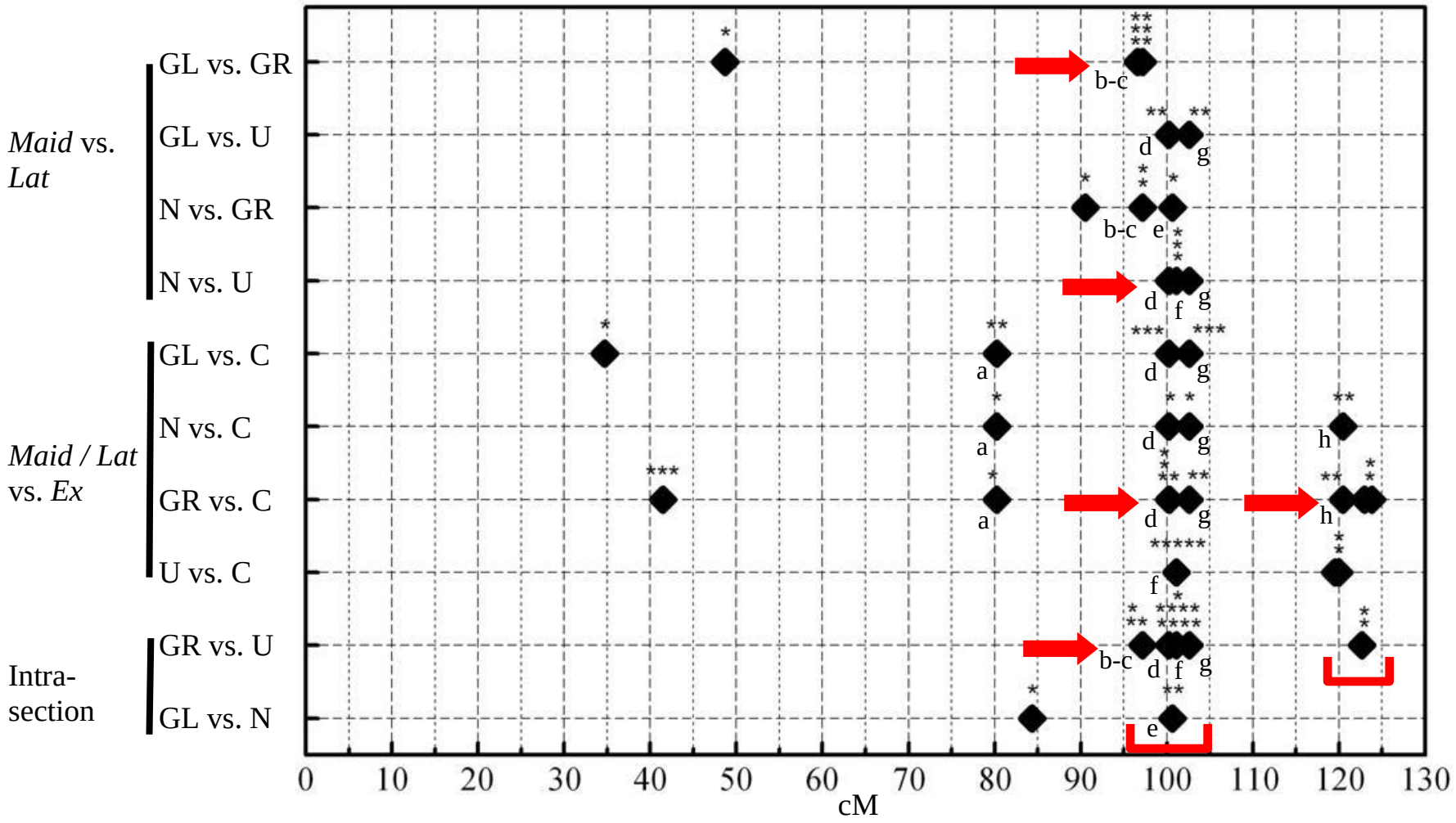

Figure 4.5K Linkage group 11; a) ePt-565082, b) ePt-567579, c)ePt-636573, d) ePt-567794, e) ePt-571164, f) ePt-503764, g) ePt-642570 and h) ePt-574631.

4.4 Discussion

4.4.1 Species differentiation and diversity

Individuals formed robust species groups with pair-wise species Φ*pt* values indicating that all species were significantly differentiated. The relative pair-wise level of species differentiation was broadly congruent with taxonomy, with species from within the same section generally being less differentiated than those from different sections. However, the inter-section comparison between *E. urophylla* (section *Latoangulatae*) and *E. camaldulensis* (section *Exsertaria,* Φ*pt* = 0.283; Table 4.3) showed marginally less differentiation than was found between *Latoangulatae* species *E. urophylla* and *E. grandis* (Φ*pt* = 0.304). Despite belonging to different sections (Brooker 2000), phylogenetic analyses have shown that *E. grandis*, *E. urophylla* and *E. camaldulensis* are closely related (Steane *et al.* 2007; Steane *et al.* 2011). Although these species have been found to consistently group within the same phylogenetic clade, analyses have shown that *E. grandis* and *E. urophylla* are more closely related to each than either *Latoangulatae* species is to *E. camaldulensis* (Steane *et al.* 2007; Steane *et al.* 2011). The slightly smaller pair-wise Φ*pt* value detected between *E. urophylla* and *E. camaldulensis* in this study may reflect historical processes. For example, it is hypothesised that *E. urophylla* diverged from a *Symphyomyrtus* taxa in Australia ~5-2 million years ago, before initially colonising the island of Timor at a time when sea levels were lower than at present (Ladiges *et al.* 2003; Payn *et al.* 2007). The closest Australian region to Timor is the northwest region of the Northern Territory in northern Australia. *Eucalyptus camaldulensis* subsp. *obtusa* is widely distributed in this area (McDonald *et al.* 2009), and based on molecular data which suggests that *E. camaldulensis* has tropical origins (Butcher *et al.* 2009), may have occupied this area for a considerable time. Therefore, if *E. urophylla* formerly occurred or originated in the same region, the relatively close inter-section affinity between these species could be explained by historical gene flow having occurred between the species in this region.

Species genetic diversity estimates were calculated using a common set of 2207 DArT markers. It is statistically difficult to estimate allele frequencies for DArT markers due to their dominant nature; *i.e.* the plus-allele (presence of a band) dominates over the null-allele (absence of a band; Zhivotovsky 1999). This can create problems when calculating genetic diversity estimates and potentially bias estimates unless species inbreeding estimates are known (Kremer *et al.* 2005). Furthermore, so-called

'ascertainment bias' associated with marker development (Clark *et al.* 2005) can also bias genetic diversity estimates. In this study a high percentage (44%) of the 2207 markers used in genetic diversity calculations were derived from section *Latoangulatae* species (*E. grandis* and *E. urophylla*) and or their hybrids (Table 4.2). Therefore, it could be expected that these markers would have a greater probability of being polymorphic in *Latoangulatae* species in comparison to species from other sections. However, both Sansaloni *et al.* (2010) and Steane *et al.* (2011) reported that the species origin of DArT markers had little influence on the probability of markers being polymorphic across species from different section. Thus, it appears that the ascertainment bias associated with DArT marker development may be low and that this may have had negligible effects on genetic diversity estimates. As range-wide species *F* estimates were not available for all of the species included in this study, both individual marker H_{HW} and H_{PH} genetic diversity estimates were calculated. An inspection of species *Q* (null-allele) frequency profiles found that all species had u-shaped *Q* distributions (Figure 4.3). Kremer *et al*. (2005) has shown that u-shaped *Q* distributions are highly robust to the bias associated with estimating species genetic diversity in the absence of *F*. In agreement with Kremer *et al*. (2005), species H_{HW} and H_{PH} estimates (mean of 2207 markers) were very similar indicating that accurate species genetic diversity estimates could be obtained using a large number of DArT markers without including *F*.

Species samples were evenly collected from throughout the full geographical range of each species. Apart from a slight over-representation of central Victorian *E. nitens* individuals, the sampling method used should not have biased genetic diversity estimates. Comparisons of genetic diversity estimates obtained in this study with previous estimates are confounded by the sampling methods of different studies in addition to the number of individuals, molecular marker type and number of markers used for analyses. Despite these issues, the relative estimates of species genetic diversity obtained in this study appear to be in accordance with previous findings. For example, species genetic diversity (assuming Hardy-Weinberg equilibrium; H_{HW}) was lowest in *E. nitens* (H_{HW} = 0.146; Table 4.4) and this likely reflects the low level of population differentiation which occurs between central Victorian populations (Byrne *et al.* 1998; Thumma *et al.* 2005) and the limited distribution of this species outside this region (Cook and Ladiges 1991). Although no previous eucalypt population genetic studies have calculated genetic diversity estimates in multiple species using a common

set of markers, two recent SNP marker development studies have provided useful insights (Grattapaglia *et al.* 2011; Kulheim *et al.* 2009). Kulheim *et al* (2009) characterised SNP variation in 23 genes involved in biosynthetic pathways leading to the formation of secondary metabolites in multiple eucalypt species. *Eucalyptus globulus* and *E. nitens* were found to have very similar levels of SNP diversity overall, reflecting their close phylogenetic relationship, whereas *E. camaldulensis* was found to have almost twice the level of SNP diversity compared to both section *Maidenaria* species (Kulheim *et al.* 2009). A similar result was obtained by Grattapaglia *et al.* (2011), who from over 1.1 million expressed sequence tags (ESTs) developed a set of 768 SNP markers for use in *Symphyomyrtus* species. SNP diversity was again found to be higher in *E. camaldulensis* relative to *Maidenaria* species (Grattapaglia *et al.* 2011). The high SNP diversity found in *E. camaldulensis* by Grattapaglia *et al.* (2011) is consistent with the relatively high level of genetic diversity detected in *E. camaldulensis* in this study. This high level of diversity probably reflects the species widespread distribution and the presence of many subspecies; suggesting much evolutionary divergence within the species as taxonomically defined (Butcher *et al.* 2009; Kulheim *et al.* 2009; McDonald *et al.* 2009). Perhaps surprising was the finding that *E. urophylla* had the greatest level of genetic diversity (H_{HW} = 0.231; Table 4.4). Although *E. urophylla* occupies a much smaller geographic range compared to *E. camaldulensis*, the species does exhibit substantial morphological variation in numerous characters and has the greatest altitudinal range of any eucalypt (reviewed in Payn *et al.* 2008); with individuals included in this study occurring from 385 to 1900 m. Therefore, considerable genetic diversity associated with localised adaptation could be expected to be found within *E. urophylla*. Indeed, Grattapaglia *et al*. (2011), found that the level of SNP diversity in *E. urophylla* was similar to *E. camaldulensis*, in agreement with the high level of genetic diversity detected in *E. urophylla* in this study.

4.4.2 Outlier marker tests

In each pair-wise species comparison, Bayescan analyses detected a total of 30 (*E. grandis* vs. *E. urophylla*) to 62 (*E. nitens* vs. *E. grandis*) outlier markers. This equated to 1.7-3.5% of the markers analysed (Table 4.5) and is consistent with the percentage of outlier markers commonly detected in genome-scan studies (typically < 5% of markers; reviewed by Holderegger *et al.* 2008). The power to detect outlier markers has been shown to decrease with increasing differentiation between populations (or species); due to the increasing background level of putatively neutral variation (Butlin 2010; Perez-

Figueroa *et al.* 2010). For example, 8.9% of markers examined were detected as outlier markers between European oak species which exhibited a low level of species differentiation (G_{ST} < 0.01; Scotti-Saintagne *et al.* 2004). In contrast, only 1.45% of markers were detected as outliers between sympatric *Howea* palm species on Lord Howe Island in which species differentiation was high (F_{ST} = 0.31; Savolainen *et al.* 2006). In this study, all species were highly differentiated (Bayescan mean marker F_{ST} = 0.26 to 0.44; Table 4.5) and this is likely to explain why only markers with extreme frequency differences (*i.e.* fixed or very high 1 ratio differences) were identified as outlier markers in many of the pair-wise species comparisons (Table 4.5). It is possible that these fixed marker differences could represent genomic regions in which hard selective sweeps have occurred; *i.e.* in which a new advantageous mutation arises and is quickly spread to fixation by natural selection (Hohenlohe *et al.* 2010; Pritchard *et al.* 2010). These 'hard sweep' signatures of selection are relatively easy to detect using genome scans. However, species adaptation can also occur through 'soft sweeps' (adaptation from standing genetic variation) and or polygenic adaptation, in which shifts in allele frequency across a large number of loci occur, each having a small but cumulative large phenotypic effect (Pritchard *et al.* 2010). These soft signatures of selection are much more difficult to detect (Pritchard *et al.* 2010). Therefore, in addition to the difficulty associated with identifying outlier markers between highly differentiated species, the number of outlier markers detected in this study is likely to underestimate the number of loci involved in the adaptive divergence of these species.

4.4.3 The 'Genomic Atlas'

The composite genetic linkage map provided a framework for the so-called 'Genomic Atlas' and the assessment of both species genetic diversity and outlier markers at a genome-wide scale. However, a number of assumptions were made by using this map and therefore results from these analyses should be cautiously interpreted. Firstly, the composite map was constructed by integrating marker information from individual linkage maps constructed in mapping families of only three of the species include in this study; *E. grandis*, *E. urophylla* and *E. globulus* (see Chapter 2 of this thesis). While comparative mapping studies have indicated that the genomes of these three species are highly syntenic and colinear (this thesis Chapter 1; Brondani *et al.* 2006; Hudson *et al.* in press; Myburg *et al.* 2003) some small genomic differences have been detected (this thesis Chapter 1; Hudson *et al.* in press). The level of genome synteny and colinearity between each of these species with both *E. nitens* and *E. camaldulensis* has been less

characterised, since only a limited number of linkage mapping studies have been conducted in these species (e.g. Agrama and Salah 2002; Byrne *et al.* 1995). However, based on the assessment of 122 common RFLP markers mapped in both *E. globulus* and *E. nitens* linkage maps, Thamurus *et al*. (2002) established high synteny and colinearity between these two *Maidenaria* species. Also, given the high level of genome synteny and colinearity detected between species from relatively distant *Symphyomyrtus* sections *Maidenaria* (e.g. *E. globulus*) and *Latoangulatae* (e.g. *E. grandis* and *E. urophylla*), it is likely that *E. camaldulensis* from section *Exsertaria* (more closely related section to *Latoangulatae*) would also have similar genome organisation (this thesis Chapter 1; Hudson *et al.* in press). Additionally, different genome sizes have also been reported for *E. urophylla*, *E. grandis* and *E. globulus*, with estimated sizes being 625, 611 and 418 Mbp, respectively (Grattapaglia and Bradshaw 1994). Therefore, although the composite map can be expected to represent the overall genome organisation of species reasonably accurately, genomic differences do exist between species and therefore localised map differences can be expected to occur.

The estimates of species genetic diversity mapped on the composite map will have been influenced somewhat by differences in marker density across the composite map; due to the use of an 11-marker running average. For example, in regions of the map with relatively low density, genetic diversity values based on an 11-marker running average are likely to have masked the true level of genetic diversity. Although an average of 128 markers were mapped on each linkage group of the composite map, many more markers, especially in regions of low marker density, would be preferable and would increase the accuracy of these estimates. Apart from making costly improvements to the genotyping array to increase the number of markers (e.g. screening additional genomic representations, decreasing the level of marker redundancy; see Sansaloni *et al.* 2010), higher marker density could have been achieved by positioning unmapped markers from the 2207 marker dataset on the composite map. As DArT marker sequences were available for all non-mapped markers (799), BLAST searches could have identified the positions of these markers on the assembled *E. grandis* genome sequence. However, due to the considerable time required to, (1) conduct BLAST searches for the many unmapped markers, and (2) use BLAST results to determine the relative positions of markers on the composite map, no attempt was made to position unmapped markers in this study. However, as the DArT markers have now been annotated on the *E. grandis* genome sequence available at EucGenIE (DArT markers annotated August 2011;

http://eucgenie.bi.up.ac.za/), which would dramatically reduce the time needed to position these unmapped markers, this may now be more feasible.

Estimates of species divergence times within subgenus *Symphyomyrtus* are uncertain. Climatic and tectonic data suggests that the radiation of species within sections may have occurred within the last 2 Mya and that divergence of sections within subgenus *Symphyomyrtus* occurred between 5-10 Mya (Ladiges *et al.* 2003). However, molecular dating indicates that these events may be older, occurring between 5-10 Mya and 13-36 Mya, respectively (Crisp *et al.* 2004). Regardless of this uncertainty, considerable time has elapsed following the divergence of species for molecular differences to accumulate within species. Thus, in the absence of biological or functional information for outlier markers, it is not possible to conclude whether markers detected as outliers have been involved in speciation (Harr 2006; Pritchard *et al.* 2010). Another important caveat to consider is that demographic processes (e.g. population bottlenecks or founder events) can result in patterns of polymorphism that mimic signatures of natural selection (Neale and Ingvarsson 2008; Oleksyk *et al.* 2010; Stinchcombe and Hoekstra 2007). Therefore, although genome-scan studies can identify signals of molecular differentiation, it is difficult to assess how much confidence can be placed in individual signals (Pritchard *et al.* 2010).

The repeated detection of outlier markers in multiple comparisons and the grouping of outlier markers in genomic hot-spots provided supporting evidence for the putative involvement of loci in species differences. Detection of the same outlier marker in multiple comparisons guards against the possibility of detecting false positives (which also increases with increasing species differentiation; see Butlin 2010). Despite only half of the unique outlier markers identified being mapped on the composite map, multiple outlier marker clusters in addition to numerous diversity deserts were identified. In particular, a number of hot-spots were detected on LGs 8 and 11 (Figure 4.5H and K) which warrant further investigation.

4.4.4 Future studies

Two broad approaches can be used to characterise and determine the functional importance of the hot-spots identified in this study. The first of these complementary approaches involves testing for the association between these genomic regions with the positions of mapped quantitative trait loci (QTL). This approach has been used successfully to examine the genetic basis of growth rate differences between dwarf and

normal ecotypes of whitefish (*Coregonus clupeaformis*; Rogers and Bernatchez 2007). Rogers and Bernatchez (2007) found that amplified fragment length polymorphism (AFLP) outlier loci co-located with QTL for growth rate; suggesting that differentiation at these loci was due to selection on nearby growth related genes. As numerous eucalypt mapping studies have mapped QTL for many diverse traits, including growth and wood properties, response to biotic and abiotic stress and chemicals involved in secondary metabolism (reviewed in Grattapaglia *et al.* submitted), the alignment of these QTL against the 'Genomic Atlas' produced in this study could provide evidence for the involvement of natural selection on the outlier markers detected.

A second approach involves the search for candidate genes. In eucalypts this can be performed by using the large transcript and EST resources which are progressively being annotated on the reference *E. grandis* genome sequence. However, while a number of hot-spots have been identified for which candidate gene searches can be performed, the lack of an explicit biological hypothesis can be both advantageous and problematic. On one hand, the genomic regions identified are likely to contain many genes (either coding or regulatory), making it difficult to identify a small number of candidates for further testing. On the other hand, genome-scans may possibly identify candidate genes which are linked to traits that have not previously been investigated using QTL analyses; thus providing novel insights into the range of adaptive traits.

Although this study was designed to investigate the molecular basis of species differences, the DArT marker data generated in this study could also be used to examine intra-specific processes. For example, analyses could be designed to provide insights into adaptation between, for example, the seven subspecies of *E. camaldulensis* (McDonald *et al.* 2009), or races within *E. globulus* (Dutkowski and Potts 1999). Previous genome scan studies have also provided a genetic framework for investigating adaptation over environmental gradients (e.g. Bonin *et al.* 2006; Jump *et al.* 2006). For example, Jump *et al.* (2006) conducted outlier tests in populations of common beech (*Fagus sylvatica*) in the Montseny mountain range of Spain; in which populations were distributed between 992-1640 m. One AFLP marker was detected to be under selection and population surveys found that this marker was fixed in the highest altitude population. Similar analyses within species, either across latitude (e.g. *E. grandis*) or altitude (e.g. *E. urophylla*), could similarly provide valuable insight into cline-based adaptations in eucalypts. From the pair-wise species fixed difference markers identified, it may also be possible to generate a series of PCR-based diagnostic markers. These

could then be used for a wide range of diagnostic tests, such as a means for identifying hybrid progeny between morphologically similar parent species when morphology-based assessment is not possible.

4.4.5 Conclusion

This pilot study confirmed that DArT can provide sufficiently large marker datasets which are shared between species and provide adequate marker resolution and genome-wide coverage for population genomic studies in eucalypts. This study was the first to, (1) estimate species genetic diversity in multiple eucalypts using a large set of common markers, and (2) investigate genome-wide species differences at the molecular level in more than two species simultaneously using population genomic methods. The mapping of genetic diversity estimates and outlier markers identified several genomic hot-spots which will be the focus of future studies. Funding is currently being sought to expand this study through the addition of more subgenus *Symphyomyrtus* species. Mapping the genomic differences detected between these additional species will importantly show whether there are consistent patterns of differentiation between species. This will help 'build the case' for conducting further tests to confirm the adaptive role of markers and or candidate genes identified in these differentiated regions. It would also be very beneficial to integrate this 'Genomic Atlas' data with other genomic resources (e.g. datasets currently annotated on the *E. grandis* genome sequence in addition to QTL position information etc.). A publicly available database containing an integrated suite of resources will greatly assist our abilities to understand the molecular basis of eucalypt adaptation and speciation.

Conclusions

The research undertaken in this thesis was conducted during a time in which sequencing, molecular and bioinformatics techniques are progressing at unprecedented rates. For example in *Eucalyptus*, the release of an assembled reference genome sequence for *E. grandis* in January 2010 (V1.0; www.phytozome.net/) represented a milestone event in eucalypt genomics. As shown in this thesis, this resource has provided new avenues for addressing questions pertaining to the development, diversity and evolution of eucalypts.

This study used Diversity Arrays Technology (DArT) molecular markers to conduct genomic analyses in five of the most commercially important eucalypt species; all from subgenus *Symphyomyrtus*. In particular, genetic linkage maps constructed with DArT markers provided the foundation for much of the research. Linkage maps are a valuable genomic tool and provide insights into the structure, organisation and evolution of plant genomes, in addition to providing a basis for QTL location. In this study, a high density linkage map was constructed in a large (n = 503) *E. globulus* dwarf x tall ecotype out-crossed F_2 family. The method used for map construction, in addition to the stringent mapping parameters applied for marker-order acceptance, and the power provided by the large progeny size of this family resulted in a high quality linkage map with robust marker-order. This 1060 DArT and microsatellite marker map is the highest density map to be produced in *E. globulus* (section *Maidenaria*). This map was used to assess synteny and colinearity with *E. grandis* (section *Latoangulatae*) through comparative mapping analyses. The high transferability of DArT markers across pedigrees and species resulted in a large number of common markers (236 – 393) between maps and provided the highest resolution yet achieved for comparative mapping in *Eucalyptus*. Despite two small putative chromosomal translocations or duplications being identified, comparative mapping indicated that the overall level of synteny and colinearity between section *Maidenaria* (*E. globulus*) and *Latoangulatae* (*E. grandis* x *E. urophylla*) was high. This finding was concordant with phylogenetic analyses which have shown that sections *Maidenaria* and *Latoangulatae* are relatively closely related within subgenus *Symphyomyrtus* (Steane *et al.* 2002; Steane *et al.* 2011). It is likely that species belonging to section *Exsertaria* (e.g. *E. camaldulensis*) will also have similar genome organisation due to their close phylogenetic relationship (Steane *et al.* 2002; Steane *et al.* 2011). Therefore, the *E. grandis* genome sequence will be highly transferable to many of the most economically important eucalypt species which belong to these three

sections (Grattapaglia and Kirst 2008; Myburg *et al.* 2007). In future studies, it will also be important to assess the level of synteny and colinearity between *E. grandis* and other commercially important species (e.g. *E. pilularis* from subgenus *Eucalyptus*) in order to similarly determine how confidently the *E. grandis* genome information can be transferred to more distantly related species.

The high synteny and colinearity detected between *E. globulus* and section *Latoangulatae* species in Chapter 1 suggested that it would be feasible to combine multiple linkage maps into a single so-called 'composite map,' while maintaining high marker-order accuracy. Several linkage maps had recently been constructed in *Symphyomyrtus* mapping pedigrees using DArT markers. Collectively, these maps contained many thousands of DArT markers and several hundred microsatellite markers. A single composite map was produced through the integration of seven independently constructed linkage maps using a marker-merging method. The composite map contained 3909 DArT, 218 microsatellite and 8 gene markers making it the highest density map produced for eucalypts. Marker-order colinearity between composite map linkage groups and component maps were mostly high, indicating that the composite map marker-orders closely represented that of the component maps and is likely to have high marker-order accuracy. The composite map will serve as a valuable reference map for eucalypt research and has already had practical application. For example, in addition to providing a framework map for the population genomic analyses conducted in Chapter 4 of this thesis, the composite map is currently being used to assist in the improvement of the current *E. grandis* genome sequence (Jenkins *pers comm.*). The first release of the assembled *E. grandis* genome sequence contained 85 Mbp (~12% of the sequenced genome) in several thousand small scaffolds which could not be anchored to the 11 main chromosome assemblies (http://web.up.ac.za/eucagen). Information from this high density linkage map can be used to assist in incorporating these unanchored scaffolds. Furthermore, the composite map will enable the relative positions of markers to be more easily determined and will facilitate higher resolution comparative analyses. This will be of particular use for comparing the positions of QTL mapped in different studies, as in many previous studies the ability to validate the expression of QTL across variable genetic backgrounds has been hampered by a lack of shared markers between the linkage maps employed for QTL detection. The inclusion of many microsatellite markers in the composite map also provides a valuable link to many eucalypt studies

conducted prior to the development of DArT markers and which share common microsatellite markers, making the information of the composite map widely applicable.

The genetic architecture underlying traits which differentiate dwarf and tall ecotypes of *E. globulus* was investigated through QTL analyses in Chapter 3. The use of highly differentiated grandparental families maximised the possibility of trait segregation in the F_2 family examined. The large progeny size enhanced the power and precision of QTL detection, while the planting of F_2 individuals at two contrasting sites enabled the stability of traits (*i.e.* genotype x environment interactions) to be examined. A total of 33 QTL were detected for the nine vegetative and reproductive phase-change, tree height and shape traits examined. The most significant QTL detected in this family was for the initiation of vegetative phase-change; a trait which was found to be highly stable across sites. This QTL explained 63% of phenotypic variation and alleles of this QTL were found to have an additive effect, strongly influencing the node at which phase-change occurred. This QTL also appeared to have pleiotropic effects on the node at which subsequent developmental phase-change processes occurred (*i.e.* development of fully adult vegetation and sexual reproduction). A micro-RNA (miRNA) gene (EgMIR156C) belonging to a family of miRNA's (Yang *et al.* 2011) known to be involved in regulating vegetative phase-change was found near the peak of this QTL and represented a strong candidate gene. The hypothesised role of this miRNA may be responsible for the heterochronic evolution of early vegetative phase-change in the dwarf ecotype of *E. globulus*. This may have been the decisive adaptation which enabled *E. globulus* individuals to colonise the exposed coastal habitat in which dwarf ecotype populations occur.

The identification of this miRNA candidate gene illustrated how the recently developed *Eucalyptus* genomic resources, in combination with results of this thesis, have provided novel insight into not only this developmental trait but eucalypt genomics in general. For example, the linkage mapping resolution provided by DArT markers resulted in several markers being mapped to a region within the 2-LOD confidence interval of the QTL detected for vegetative phase-change. Having shown that the genomes of *E. globulus* and *E. grandis* are highly colinear (Chapter 1), the *E. grandis* genome sequence within this region could be found with high confidence using the available DArT marker sequences. This subsequently led to the identification of the EgMIR156C candidate gene which will be the focus of future studies.

In the final chapter of this thesis (Chapter 4), a so-called ‘Genomic Atlas’ was produced for five ecologically diverse and economically important species from subgenus *Symphyomyrtus*. By plotting genetic diversity estimates and markers detected as being under putative directional selection on the composite map produced in Chapter 2, several genomic ‘hot-spots’ were identified. The multiple pair-wise species comparisons conducted and the repeated detection of the same marker and/or the clustering of outlier markers provided evidence for the loci residing in these hot-spot regions being putatively involved in species differentiation. Despite being a pilot study, the results obtained were encouraging. The addition of more subgenus *Symphyomyrtus* species to this project, a search for candidate genes in the identified hot-spot regions, and the alignment of QTL positions against this atlas, provides the potential to further our understanding of adaptation and speciation between these ecologically and commercially important species.

Supplementary material

Supplementary files

File S1.1 Microsatellite screening in the *E. globulus* F_2 Lighthouse mapping family
Before genotyping the *E. globulus* F_2 Lighthouse mapping population (n = 503) with 50+ microsatellite markers and an *Eucalyptus* DArT genotyping array, a pedigree check was undertaken at the University of Tasmania (UTAS) using seven microsatellite markers (UTAS screening coloads Table S1.2). A second stage of microsatellite genotyping was undertaken at EMBRAPA (The Brazilian Agricultural Research Corporation), Genetic Resources and Biotechnology Laboratories, Estação Parque Biológico, Brasília DF, Brazil. This involved screening 210 EMBRA and 35 CSIRO microsatellite markers (Table S1.2) for amplification quality and degree of polymorphism. This screening was conducted in both parents of the F_2 family (individuals BA0010 and BA0012) and a subset of 10 F_2 individuals. Microsatellite markers were amplified in multiplex PCR reactions as detailed in the materials and methods (Chapter 1) and scored as either Failed (did-not amplify), Poor (either a weak or complex pattern of amplification was produced and could-not be confidently scored), or Good (see column 'Amp' in Table S1.2 and Table S1.3). No PCR optimisation was attempted for microsatellite markers which either failed or produced poor amplification products.

Following microsatellite screening, 48 microsatellite markers were chosen for genotyping the entire F_2 family. These were chosen based on; (1) marker amplification quality and degree of polymorphism (Table S1.3), (2) genome distribution of microsatellite markers; with positions being estimated from previously published (Brondani *et al.* 2006; Freeman *et al.* 2006; Thamarus *et al.* 2002) and unpublished (D. A. Faria and L. Neves; both EMBRAPA Genetic Resources and Biotechnology, Brazil) linkage maps, (3) if the microsatellite had been genotyped in the *E. grandis* x *E. urophylla* double pseudo-backcross mapping pedigree (Kullan *et al*., in press), and (4) whether the microsatellite had shown QTL association with phenotypic traits similar those to be investigated in this mapping family in previous studies (e.g. floral and vegetative phase change traits; Freeman 2006; Missiagia *et al.* 2005; Thamarus *et al.* 2002). The final 48 microsatellite markers were amplified in nine multiplexes, with new multiplex PCR configurations (Genotyping coloads; Table S1.2) firstly tested in 16 F_2 individuals for amplification quality prior to genotyping the entire F_2 family.

Supplementary tables

Table S1.1 Summary of 'additional' *E. globulus* integrated consensus linkage maps (F_1 FAM1, F_1 FAM5 and KI x T F_2) which were used to investigate the linkage group position of non-syntenic markers.

Map and LG	cM	Marker interval cM		Markers mapped					
		Av.	Max	DArT	SSR	Gene	AFLP	Total	Seg. dist. (%)[a]
***E. globulus* F_1 FAM1**									
1	109.9	2.56	16.9	52	1			53	0 (0)
2	114.1	2.72	22.3	52				52	0 (0)
3	114.1	1.78	11.6	78		1		79	14 (17.7)
4	80.9	2.70	23.6	40				40	0 (0)
5	101.8	1.27	11.9	101				101	3 (3)
6	84.3	2.41	25.5	43	1			44	9 (20.5)
7	82.3	1.58	11.1	55		1		56	2 (3.6)
8	97.7	1.53	13.8	81	1			82	18 (22)
9	76.8	1.87	11.8	55				55	0 (0)
10	66.1	2.13	15.6	36				36	5 (13.9)
11	105.6	2.51	16.7	57	2			59	2 (3.4)
Total	1033.5			650	5	2	0	657	53
Average		1.97		59.1				59.7	8.1%
***E. globulus* F_1 FAM5**									
1	111.0	3.3	23.8	45	1			46	6 (13.0)
2	69.3	1.5	20.7	52				52	2 (3.8)
3	110.7	1.5	12.7	88		1		89	25 (28.1)
4	83.3	3.8	17.8	29				29	0 (0)
5	104.0	1.9	22.7	66				66	1 (1.5)
6	106.0	3.7	14.0	40	1			41	2 (4.9)
7	85.4	1.8	19.8	69				69	0 (0)
8	107.7	1.6	16.1	79	1			80	36 (45.0)
9	78.9	2.7	10.9	38				38	1 (2.6)
10	86.7	1.6	14.2	69		1		70	5 (7.1)
11	112.8	2.4	19.1	54	1			55	3 (5.5)
Total	1055.9			629	4	2	0	635	81
Average		2.1		57.2				57.7	12.8%
***E. globulus* KI x T F_2**									
1	96.2	2.0	14.3	42	6		9	57	0 (0)
2	124.0	3.0	23.6	46	4		9	59	27 (45.8)
3	121.3	1.7	18.4	81	2	1	10	94	10 (10.6)
4	83.6	2.5	16.1	46	4	2	9	61	3 (4.9)
5	109.9	1.8	8.7	71	3		17	91	25 (27.5)
6	151.5	3.1	26.4	50	4		12	66	28 (42.4)
7	112.4	1.7	9.5	74	4	1	14	93	4 (4.3)
8	147.2	2.1	12.1	85	3		12	100	10 (10)
9	90.6	2.5	14.9	59	1		5	65	6 (9.2)
10	102.5	2.3	17.0	61	2	1	4	68	16 (23.5)
11	123.2	2.9	15.2	45	3		8	56	5 (8.9)
Total	1262.5			660	36	5	109	810	134
Average		2.2		60.0	3.3		9.9	73.6	16.5%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$.

Table S1.2 Microsatellite markers screened and genotyped in the *E. globulus* F_2 Lighthouse family. Multiplex co-load, allele size range, degree of polymorphism (No. alleles and seg.= segregation type) and amplification quality (Amp. = Amplification score; 0=Failed, 1=Poor and 2=Good) are shown for each microsatellite marker.

Coload	Primer type	Number and ref[J]	Amp	Genotyped microsatellite coload[K]	Allele size range	No. alleles	Seg[L]
Br_1	EMBRA	11[D]	2		109-128	2	4
Br_1	EMBRA	70[E]	2		143-170	3	2
Br_1	EMBRA	147[F]	0				
Br_1	EMBRA	180[F]	2		123-128	3	2
Br_1	EMBRA	333[F]	1		209-227		
Br_1	EMBRA	676[H]	1		257-259		
Br_10	EMBRA	22[E]	0				
Br_10	EMBRA	41[E]	2	3	175-205	4	1
Br_10	EMBRA	53[E]	2	3	108-126	3	4
Br_10	EMBRA	196[F]	2		256-277	4	1
Br_10	EMBRA	269[F]	2		190-202	3	3
Br_10	EMBRA	310[F]	2		269-275	3	4
Br_11	EMBRA	3[D]	2		114-143	3	4
Br_11	EMBRA	37[E]	2	4	129-157	3	2
Br_11	EMBRA	42[E]	2		156-183	3	2
Br_11	EMBRA	153[F]	2	4	194-200	3	2
Br_11	EMBRA	210[F]	2		205-227	4	1
Br_11	EMBRA	345[F]	2		218-232	4	1
Br_12	Eg	91[C]	2		134-155	4	1
Br_12	EMBRA	80[F]	2		061-079	2	5
Br_12	EMBRA	121[F]	2	1	112-140	4	1
Br_12	EMBRA	173[F]	1		080-105		
Br_12	EMBRA	696[H]	0				
Br_12	EMBRA	746[H]	2	9	179-215	3	2
Br_13	EMBRA	78[F]	2	7	120-129	2	4
Br_13	EMBRA	87[F]	2		228-267	4	1
Br_13	EMBRA	135[F]	0				
Br_13	EMBRA	187[F]	0				
Br_13	EMBRA	388[F]	1		113-141		
Br_13	EMBRA	747[H]	2	9	264-284	4	1
Br_14	EMBRA	72[F]	2	5	132-145	3	2
Br_14	EMBRA	105[F]	0				
Br_14	EMBRA	148[F]	2		186-212	4	1
Br_14	EMBRA	361[F]	0				
Br_14	EMBRA	681[H]	2		225-235	2	5
Br_14	EMBRA	691[H]	2		211-229	4	1
Br_15	EMBRA	51[E]	0				
Br_15	EMBRA	120[F]	2		131-149	4	1
Br_15	EMBRA	179[F]	2		125-127	2	5
Br_15	EMBRA	674[H]	1		184-184		
Br_15	En	6[I]	2		086-093	3	2
Br_16	EMBRA	40[E]	2		109-109	1	
Br_16	EMBRA	64[E]	1		160-160		
Br_16	EMBRA	125[F]	1		143-154		
Br_16	EMBRA	131[F]	2		086-090	2	3
Br_16	EMBRA	201[F]	2		158-162	3	2
Br_16	EMBRA	203[F]	1		323-345		
Br_17	EMBRA	69[E]	1		101-119		
Br_17	EMBRA	91[F]	1		118-138		
Br_17	EMBRA	100[F]	0				
Br_17	EMBRA	112[F]	0				
Br_17	EMBRA	286[F]	2		138-154	2	6
Br_17	Es	140[A]	2		132-137	2	6
Br_18	EMBRA	5[D]	0				

Br_18	EMBRA	47[E]	2	8	115-128	3	2
Br_18	EMBRA	157[F]	2		127-127	1	
Br_18	EMBRA	186[F]	0				
Br_18	EMBRA	945[H]	1		087-094		
Br_18	EMBRA	979[H]	1		405-415		
Br_19	EMBRA	48[E]	1		135-144		
Br_19	EMBRA	324[F]	2		141-141	1	
Br_19	EMBRA	695[H]	2		112-114	2	5
Br_19	EMBRA	915[G]	2		210-210	1	
Br_19	EMBRA	1627[H]	2		094-094	1	
Br_19	EMBRA	1639[H]	2		179-179	1	
Br_2	Eg	96[C]	2	1	273-279	2	4
Br_2	EMBRA	58[E]	2	1	134-167	3	2
Br_2	EMBRA	68[E]	1		085-120		
Br_2	EMBRA	111[F]	2		119-147	2	5
Br_2	EMBRA	632[H]	0				
Br_2	EMBRA	705[H]	2		263-265	2	4
Br_20	EMBRA	7[D]	2		134-167	3	2
Br_20	EMBRA	127[F]	2	3	141-149	4	1
Br_20	EMBRA	209[F]	0				
Br_20	EMBRA	242[F]	0				
Br_20	EMBRA	331[F]	0				
Br_20	EMBRA	2000[H]	2		237-239	2	4
Br_21	Eg	128[C]	1		182-215		
Br_21	EMBRA	10[D]	1		128-145		
Br_21	EMBRA	155[F]	2		143-160	3	4
Br_21	EMBRA	1474[H]	0				
Br_21	EMBRA	1656[H]	2	2	102-109	3	2
Br_21	EMBRA	1924[H]	1		324-330		
Br_22	EMBRA	21[E]	0				
Br_22	EMBRA	81[F]	1		082-121		
Br_22	EMBRA	165[F]	1		099-111		
Br_22	EMBRA	214[F]	1		117-145		
Br_22	EMBRA	383[F]	2		124-162	3	4
Br_22	EMBRA	1071[H]	2		268-270	2	5
Br_23	EMBRA	31[E]	1		154-154		
Br_23	EMBRA	250[F]	0				
Br_23	EMBRA	1144[H]	2		154-154	1	
Br_23	EMBRA	1507[H]	2		115-120	2	4
Br_23	EMBRA	1793[H]	2		210-218	3	2
Br_23	En	16[I]	2	5	165-184	4	1
Br_24	EMBRA	168[F]	1		076-078		
Br_24	EMBRA	357[F]	0				
Br_24	EMBRA	950[H]	2		181-181	1	
Br_24	EMBRA	1953[H]	1		198-199		
Br_24	EMBRA	2055[H]	2		148-148	1	
Br_3	EMBRA	12[D]	2	1	126-135	3	2
Br_3	EMBRA	122[F]	2		208-208	1	
Br_3	EMBRA	219[F]	2	1	257-273	3	2
Br_3	EMBRA	350[F]	2		202-225	2	5
Br_3	EMBRA	390[F]	2		131-144	3	2
Br_3	EMBRA	648[H]	2		165-165	1	
Br_4	Eg	98[C]	1		177-181		
Br_4	EMBRA	43[E]	1		091-107		
Br_4	EMBRA	63[E]	2	5	176-184	3	2
Br_4	EMBRA	172[F]	2		279-301	4	1
Br_4	EMBRA	367[F]	1		073-087		
Br_5	EMBRA	98[F]	2	3	221-241	3	2
Br_5	EMBRA	106[F]	0				
Br_5	EMBRA	126[F]	0				
Br_5	EMBRA	290[F]	0				
Br_5	EMBRA	332[F]	1		355-369		

Br_5	EMBRA	645[H]	0				
Br_6	EMBRA	27[E]	2	2	114-127	3	2
Br_6	EMBRA	32[E]	0				
Br_6	EMBRA	213[F]	2		207-215	2	5
Br_6	EMBRA	618[H]	2	2	149-151	2	3
Br_6	EMBRA	623[H]	2		204-227	4	1
Br_6	EMBRA	646[H]	2		139-158	3	2
Br_7	Eg	62[C]	2		197-197	1	
Br_7	EMBRA	9[D]	1		114-121		
Br_7	EMBRA	28[E]	2	3	200-206	3	2
Br_7	EMBRA	128[F]	1		123-139		
Br_7	EMBRA	208[F]	1		075-114		
Br_7	EMBRA	239[F]	1		156-189		
Br_8	EMBRA	4[D]	1		080-082		
Br_8	EMBRA	18[D]	2	9	120-129	3	2
Br_8	EMBRA	45[E]	1		126-137		
Br_8	EMBRA	61[E]	1		162-172		
Br_8	EMBRA	94[F]	1		217-239		
Br_8	EMBRA	668[H]	1		186-193		
Br_9	EMBRA	8[D]	2		131-152	4	1
Br_9	EMBRA	36[E]	2	2	166-185	3	2
Br_9	EMBRA	202[F]	1		220-226		
Br_9	EMBRA	627[H]	2	2	226-237	3	3
Br_9	EMBRA	629[H]	2		179-181	2	3
Br_9	EMBRA	731[H]	2		244-249	2	4
Br_C01	Eg	99[C]	2	6	190-199	3	2
Br_C01	EMBRA	124[F]	0				
Br_C01	EMBRA	146[F]	1		087-093		
Br_C01	EMBRA	204[F]	2	6	137-161	4	1
Br_C01	EMBRA	266[F]	0				
Br_C01	En	13[C]	2		083-088	3	2
Br_C02	Eg	128[C]	1		181-214		
Br_C02	EMBRA	119[F]	2		188-202	3	2
Br_C02	EMBRA	167[F]	0				
Br_C02	EMBRA	176[F]	0				
Br_C02	EMBRA	222[F]	2	5	069-081	4	1
Br_C02	Es	115[A]	1		096-120		
Br_C03	Eg	24[C]	2		171-171	1	
Br_C03	EMBRA	60[E]	1		082-096		
Br_C03	EMBRA	88[F]	1		087-094		
Br_C03	EMBRA	158[F]	2		096-119	3	3
Br_C03	EMBRA	184[F]	0				
Br_C04	Eg	89[C]	2		164-167	2	4
Br_C04	EMBRA	164[F]	1		112-122		
Br_C04	EMBRA	227[F]	2	6	306-315	3	2
Br_C04	Es	54[A]	2	4	096-120	3	2
Br_C05	Eg	86[C]	2		201-208	3	2
Br_C05	EMBRA	233[F]	1		160-172		
Br_C05	EMBRA	240[F]	2	8	319-345	3	2
Br_C05	Es	157[A]	2	2	112-114	3	2
Br_C06	Eg	94[C]	1		094-119		
Br_C06	EMBRA	19[D]	2		162-162	1	
Br_C06	Es	140[A]	1				
Br_C07	Eg	67[C]	1		176-194		
Br_C07	Eg	76[C]	1		113-135		
Br_C07	EMBRA	191[F]	1		183-183		
Br_C08	Eg	111[F]	2	6	265-309	4	1
Br_C08	En	11[I]	1		126-151		
Br_C09	Eg	23[C]	2	8	267-271	3	2
Br_C09	Eg	84[C]	2	9	126-151	4	1
Br_C09	Eg	115[C]	2		186-210	3	2
Br_C10	Eg	16[C]	2		233-241	2	4

Br_C10	En	14[I]	2	6	137-149	4	1
Br_C11	Eg	30[C]	2		290-290	1	
Br_C11	EMBRA	15[D]	1		088-106		
Br_C11	EMBRA	56[E]	1		115-154		
Br_C12	Eg	65[C]	2	7	243-264	4	1
Br_C12	EMBRA	16[D]	2		118-150	4	1
Br_C13	Eg	61[C]	2		327-334	3	2
Br_C13	En	15[C]	2		080-090	2	5
Br_C13	Es	76[A]	2	4	132-180	3	2
Br_C14	Eg	117[C]	2		177-186	3	2
Br_C14	EMBRA	13[D]	1		080-080		
Br_C15	EMBRA	6[D]	1		120-139		
Br_C15	En	10[C]	1		146-150		
Br_C16	EMBRA	49[E]	1		117-127		
Br_S2_1	EMBRA	108[F]	2		161-169	2	5
Br_S2_1	EMBRA	132[F]	1		338-338		
Br_S2_1	EMBRA	168[F]	2		076-078	2	4
Br_S2_1	EMBRA	662[H]	2		282-315	3	2
Br_S2_10	EMBRA	97[F]	0				
Br_S2_10	EMBRA	302[F]	1		157-163		
Br_S2_10	EMBRA	357[F]	1		088-111		
Br_S2_11	EMBRA	152[F]	1		083-099		
Br_S2_11	EMBRA	237[F]	0				
Br_S2_12	EMBRA	160[F]	1		070-080		
Br_S2_12	EMBRA	1193[H]	1		130-140		
Br_S2_13	EMBRA	393[F]	1		124-135		
Br_S2_14	EMBRA	101[F]	0				
Br_S2_14	EMBRA	692[H]	0				
Br_S2_15	EMBRA	165[F]	1		99-111		
Br_S2_15	EMBRA	1063[H]	2	7	209-212	4	1
Br_S2_16	EMBRA	1722[H]	2		151-157	2	4
Br_S2_2	EMBRA	44[E]	2	7	194-203	3	2
Br_S2_2	EMBRA	77[E]	0				
Br_S2_2	EMBRA	224[F]	1		173-194		
Br_S2_2	EMBRA	385[F]	2	2	174-200	3	2
Br_S2_2	EMBRA	922[H]	1		094-102		
Br_S2_2	EMBRA	2014[G]	2		112-123	2	5
Br_S2_3	EMBRA	20[D]	2	7	130-139	3	3
Br_S2_3	EMBRA	171[F]	0				
Br_S2_3	EMBRA	205[F]	0				
Br_S2_3	EMBRA	1643[H]	2	8	341-354	4	1
Br_S2_3	EMBRA	1761[H]	2	8	189-191	2	4
Br_S2_4	EMBRA	29[D]	1		235-273		
Br_S2_4	EMBRA	38[E]	0				
Br_S2_4	EMBRA	651[F]	1		96-103		
Br_S2_4	EMBRA	1428[H]	2		177-189	2	3
Br_S2_4	EMBRA	1492[H]	2		333-337	3	4
Br_S2_5	EMBRA	46[E]	0				
Br_S2_5	EMBRA	362[F]	2		112-134	3	2
Br_S2_5	EMBRA	917[H]	2		200-203	2	4
Br_S2_5	EMBRA	1770[H]	2		109-111	2	5
Br_S2_5	EMBRA	1928[H]	1		318-330		
Br_S2_6	EMBRA	214[F]	1		114-145		
Br_S2_6	EMBRA	941[H]	2	4	222-241	4	1
Br_S2_6	EMBRA	1319[H]	2	7	178-181	4	1
Br_S2_7	EMBRA	39[E]	2		120-124	2	4
Br_S2_7	EMBRA	143[F]	1		166-192		
Br_S2_7	EMBRA	347[F]	1		182-184		
Br_S2_7	EMBRA	669[H]	2	6	095-099	3	2
Br_S2_7	EMBRA	943[G]	1		440-460		
Br_S2_8	EMBRA	33[E]	2		118-118	1	
Br_S2_8	EMBRA	115[F]	2	8	101-108	4	1

Br_S2_8	EMBRA	144[F]	2	209-214	2	5
Br_S2_8	EMBRA	665[H]	2	134-142	3	2
Br_S2_9	EMBRA	2[D]	2	139-143	2	4
Br_S2_9	EMBRA	138[F]	1	182-202		
Br_S2_9	EMBRA	189[F]	2	139-143	3	2
UTAS_A	CRC	6[B]	2	155-185	4	1
UTAS_A	EMBRA	173[F]	2	079-103	4	1
UTAS_A	En	12[C]	2	209-237	3	2
UTAS_B	CRC	10[B]	2	311-333	4	1
UTAS_B	CRC	11[B]	2	227-241	4	1
UTAS_B	EMBRA	11[D]	2	110-128	2	4
UTAS_B	EMBRA	30[E]	2	103-121	4	1

[J]Microsatellite number and reference; [A]Glaubitz *et al.*(2001), [B]Steane *et al.* (2001), [C] http://www.ffp.csiro.au/tigr/molecular/eucmsps.html, [D]Brondani *et al.* (1998), [E]Brondani *et al.* (2002), [F]Brondani *et al.* (2006), [G]Faria *et al*. (2011) [H]Embrapa unpublished and [I]Byrne *et al.* (1996). [K]Final multiplex coload of the 48 microsatellites used to genotype the entire F_2 family at Embrapa (Brazil) following microsatellite screening. [L]Segregation (Seg) of microsatellite alleles; 1=ABxCD, 2=EFxEG, 3=LMxLL, 4=NNxNP, 5=ABxAB and 6=AAxBB.

Table S1.3 Amplification quality and polymorphism of 210 Embra and 35 CSRIO (En, Es and Eg) microsatellites (SSRs) screened in the *E. globulus* F_2 Lighthouse family.

Screened SSRs		Amplification			Segregation of 'Good' SSRs			
		Failed	Poor	Good	Uninformative	From female parent	From male parent	[a]From both parents
Embra	210	39	64	107	26	8	18	55
En	7	0	2	5	1	0	0	4
Es	6	0	2	4	1	0	0	3
Eg	22	0	6	16	3	0	3	10
Total	245	39	75	132	31	8	21	72

[a]3 or 4 alleles segregating.

Table S1.4 Summary of *E. grandis* and *E. urophylla* parental linkage maps and the *E. grandis* x *E. urophylla* consensus (GU) linkage map.

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	Seg. dist. (%)[a]
***E. grandis* map**							
1	89.9	1.19	17.7	70	5	75	16 (21.3)
2	92.3	0.84	9.3	103	6	109	25 (22.9)
3	71.6	0.68	6.7	103	4	107	35 (32.7)
4	69.4	0.90	6.8	75	2	77	23 (29.8)
5	82.0	0.75	10.6	107	2	109	43 (39.4)
6	90.5	0.78	7.3	102	3	105	26 (24.7)
7	82.0	0.93	11.6	86	3	89	23 (25.8)
8	98.7	0.81	8.6	118	3	121	13 (15.7)
9	79.6	0.95	9.1	80	3	83	30 (36.1)
10	82.5	1.96	9.8	41	1	42	17 (40.4)
11	86.2	1.16	8.1	72	2	74	19 (25.6)
Total	924.7			957	34	991	270
Average		0.99		87	3.1	90.0	27.5%
***E. urophylla* map**							
1	85.1	1.27	8.7	61	6	67	13 (19.4)
2	107.1	1.00	11.4	101	6	107	27 (25.2)
3	113.4	1.09	6.7	99	5	104	40 (38.4)
4	87.4	1.50	7.9	54	4	58	12 (20.6)
5	98.5	1.09	13.1	86	4	90	60 (66.6)
6	124.8	1.23	8.9	98	3	101	25 (24.7)
7	87.1	1.04	7.8	78	5	83	56 (67.4)
8	123.1	0.93	9.8	126	6	132	69 (52.2)
9	79.0	1.25	11.2	59	4	63	12 (19.0)
10	93.7	1.51	11.2	61	1	62	14 (22.5)
11	108.2	1.28	8.4	89	2	91	20 (21.9)
Total	1107.3			912	46	956	348
Average		1.19		82.9	4.2	86.9	36.3%
***E. grandis* x *E. urophylla* consensus (GU)**							
1	88.8	0.49	4.4	173	7	180	39 (21.6)
2	102.1	0.43	5.0	228	7	235	46 (19.5)
3	105.5	0.41	3.1	251	6	257	85 (33.0)
4	79.8	0.48	5.5	157	6	163	25 (15.3)
5	110.4	0.50	9.6	218	2	220	104 (47.2)
6	136.9	0.58	8.2	232	4	236	60 (25.4)
7	83.5	0.49	7.0	163	6	169	68 (40.2)
8	119.1	0.45	6.6	263	11	274	61 (22.2)
9	88.5	0.42	6.5	203	7	210	31 (14.7)
10	97.7	0.62	5.0	155	1	156	44 (28.2)
11	95.3	0.50	4.8	186	4	190	47 (24.7)
Total	1107.6			2229	61	2290	610
Average		0.48		202.6	5.5	208.1	26.6%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$.

Table S1.5 Summary of *E. globulus* F_2 Lighthouse comprehensive parental and integrated consensus linkage maps.

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	Seg. dist. (%)[a]
Female map							
1	96.8	2.42	14.15	36	5	41	9 (22)
2	129.7	1.44	16.07	86	5	91	3 (3.3)
3	103.6	1.02	13.46	100	3	103	17 (16.5)
4	81.1	1.69	10.35	46	3	49	9 (18.4)
5	98.4	1.30	15.50	72	5	77	18 (23.4)
6	140.6	1.67	16.82	77	8	85	10 (11.8)
7	96.2	1.08	22.09	87	3	90	24 (26.7)
8	125.0	1.37	14.78	88	4	92	8 (8.7)
9	84.3	1.26	13.07	65	3	68	24 (35.3)
10	97.1	1.67	18.31	55	4	59	3 (5.1)
11	94.6	2.96	15.66	30	3	33	3 (9.1)
Total	1147.5			742	46	788	128
Average		1.48		67.5	4.2	71.6	16.2%
Male map							
1	84.5	1.80	13.04	42	6	48	11 (22.9)
2	113.2	1.40	13.90	77	5	82	2 (2.4)
3	106.3	1.18	9.00	88	3	91	19 (20.9)
4	82.8	1.53	11.64	52	3	55	11 (20)
5	101.5	1.49	18.97	66	3	69	12 (17.4)
6	139.6	1.89	16.18	70	5	75	9 (12)
7	78.3	0.97	14.10	77	5	82	29 (35.4)
8	130.1	1.57	18.68	79	5	84	8 (9.5)
9	84.1	1.24	12.99	66	3	69	32 (46.4)
10	100.7	1.98	27.00	49	3	52	3 (5.8)
11	107.4	2.90	14.08	33	5	38	5 (13.2)
Total	1128.7			699	46	745	141
Average		1.54		63.5	4.2	67.7	18.9%
Integrated consensus map							
1	94.5	1.55	13.51	56	6	62	12 (19)
2	123.3	1.00	8.86	118	6	124	5 (4)
3	102.7	0.91	9.08	112	2	114	24 (21)
4	79.0	1.36	11.09	57	2	59	7 (12)
5	101.5	0.85	12.16	115	5	120	28 (23)
6	141.4	1.24	13.01	107	8	115	17 (15)
7	91.7	0.82	11.54	109	4	113	35 (31)
8	130.2	1.10	13.82	114	5	119	10 (8)
9	85.3	0.92	11.72	91	3	94	33 (35)
10	98.1	1.23	18.71	77	4	81	3 (4)
11	103.4	1.78	12.07	54	5	59	8 (14)
Total	1151.3			1010	50	1060	182
Average		1.10		91.8	4.5	96.4	17.2%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$.

Table S1.6 Summary of *E. globulus* F_1 FAM4 parental and integrated consensus linkage maps.

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	Seg. dist. (%)[a]
Female map							
1	40.3	2.52	19.94	16	1	17	17 (0)
2	69.0	2.38	17.47	30		30	30 (13.3)
2B	35.9	3.26	15.68	12		12	12 (0)
3	128.0	2.78	32.15	47		47	47 (10.6)
4	80.7	3.23	24.03	25	1	26	26 (19.2)
5	111.1	3.37	23.35	34		34	34 (0)
6	33.0	2.36	7.30	13	1	15	15 (6.7)
6B	13.6	3.40	8.97	5		5	5 (0)
7	69.6	4.09	20.60	18		18	18 (0)
8	97.2	1.52	23.22	64	1	65	65 (0)
9	91.6	3.39	45.20	28		28	28 (17.9)
10	107.8	8.29	27.58	14		14	14 (0)
11	98.7	3.08	18.08	32	1	33	33 (27.3)
Total	976.4			338	5	344	29
Average		2.95		26		26	8.40%
Male map							
1	87.7	3.02	18.94	28	2	30	0 (0)
2	108.1	3.00	20.69	37		37	6 (16.2)
3	106.2	2.79	26.58	38		39	0 (0)
4	29.7	2.47	10.68	13		13	6 (46.2)
5	91.0	1.86	18.71	50		50	25 (50)
6	67.2	4.48	21.36	14	1	16	0 (0)
7	90.6	3.12	21.25	29		30	0 (0)
8	90.5	2.66	22.24	34	1	35	0 (0)
9	76.4	2.64	23.40	30		30	21 (70)
10	80.6	2.24	13.21	36		37	0 (0)
11	98.7	2.74	21.37	35	2	37	1 (2.7)
Total	926.5			344	6	354	59
Average		2.70		31		32	16.70%
Integrated consensus map							
1	91.2	2.61	19.97	34	2	36	0 (0)
2	111.2	1.79	15.97	63		63	9 (14.3)
3	125.6	1.63	10.02	77		78	5 (6.4)
4	97.6	3.05	24.10	32	1	33	6 (18.2)
5	117.1	1.56	19.09	76		76	25 (32.9)
6	86.6	2.99	21.28	28	1	30	1 (3.3)
7	90.8	2.16	13.14	42		43	0 (0)
8	97.8	1.24	11.63	79	1	80	0 (0)
9	110.0	2.97	23.40	38		38	21 (55.3)
10	112.7	2.40	21.25	47		48	0 (0)
11	97.5	2.27	22.83	42	2	44	9 (20.5)
Total	1137.9			558	7	569	76
Average		2.04		50.8		51.8	13.40%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$.

Table S1.7 NCBI BLAST (blseq2) sequence identity and e-values for non-syntenic marker-pairs mapped to the same positions within each consensus map in the *E. grandis* x *E. urophylla* (GU) consensus *vs. E. globulus* F_2 Lighthouse (*glob* F_2 LH) comparative mapping analysis.

		Linkage map position					Blast sequence similarity	
		GU consensus		glob F_2 LH		# base		
Pair	Marker	LG	cM	LG	cM	pairs[a]	Identity	e-value
1	ePt-636589[b]	3	82.9	7	47.5	971	309/317	7e-109
	ePt-638853	3	82.8	7	47.5	297		
2	ePt-568865	5	35.2	6	138.9	483	483/485	0
	ePt-572057	5	35.3	6	138.9	483		
3	ePt-599965	5	12.4	3	53.7	820	781/850 (gaps 33/850)	0
	ePt-643259	5	12.4	3	53.7	845		
4	ePt-574367	5	58.9	3	2.5	434	350/378 (gaps 2/378)	1e-158
	ePt-643036	5	59.0	3	2.5	379		

[a]Marker sequence length with adapter sequence/cloning vector fragments removed. [b]Marker ePt-636589 was a poor sequence, the raw sequence length for this marker is shown.

Table S2.1 Details of multicopy DArT markers mapped in the *Eucalyptus* multi-species composite map. The linkage group (LG) and position (cM) is shown for each multicopy marker position.

Multicopy marker name and the number of LGs to which it was mapped	LG	cM	LG	cM	LG	cM	LG	cM
Markers mapped to two linkage groups								
ePt-502845	4	45.4	10	81.1				
ePt-503517	4	54.9	6	47.8				
ePt-503609	9	79.6	11	48.3				
ePt-503782	8	93.8	11	116.9				
ePt-504105	4	30.9	7	48.9				
ePt-504147	5	43.2	10	23.7				
ePt-504766	5	69.3	10	40.2				
ePt-504858	5	42.4	7	13.6				
ePt-504882	2	78.8	11	51.2				
ePt-563487	2	3.5	11	44.6				
ePt-564413	6	41.2	8	116.2				
ePt-564765	6	0.0	8	90.0				
ePt-564874	4	26.2	8	106.9				
ePt-565718	6	1.8	8	88.9				
ePt-565956	2	3.3	11	44.6				
ePt-566051	7	97.1	11	29.7				
ePt-566325	5	95.9	7	74.1				
ePt-566596	2	79.3	11	48.8				
ePt-566850	2	34.7	11	39.8				
ePt-567151	9	49.3	10	32.3				
ePt-567610	8	19.1	9	23.7				

ePt-568036	2	72.2	4	11.7
ePt-568705	4	31.2	7	48.6
ePt-568743	5	69.6	6	45.6
ePt-568865	5	35.2	6	121.1
ePt-569681	2	62.9	5	72.4
ePt-569828	2	48.3	7	73.9
ePt-570269	3	48.0	10	83.1
ePt-570278	3	47.1	9	21.0
ePt-570585	3	16.1	9	93.9
ePt-570701	3	47.1	9	19.5
ePt-571374	7	97.1	11	30.4
ePt-571507	5	58.9	6	112.2
ePt-571769	4	24.5	6	6.4
ePt-571799	3	12.7	6	37.7
ePt-571831	3	85.7	5	46.7
ePt-572057	5	35.3	6	121.2
ePt-572821	2	2.2	11	44.6
ePt-573059	7	18.8	11	29.8
ePt-573788	1	33.9	7	84.0
ePt-573955	3	78.2	11	53.4
ePt-574289	5	73.5	7	71.0
ePt-574798	10	113.8	11	49.7
ePt-575049	1	74.0	11	77.8
ePt-575450	2	57.1	10	76.9
ePt-575597	5	72.6	9	28.7
ePt-599923	3	42.4	7	79.5
ePt-599965	3	55.2	5	12.4
ePt-600068	2	26.2	7	48.8
ePt-636534	1	63.8	5	60.4
ePt-636589	3	88.3	7	51.0
ePt-636681	2	43.6	3	85.5
ePt-636919	2	73.9	9	0.0
ePt-637034	4	48.7	10	81.9
ePt-637503	4	28.3	7	48.8
ePt-637656	1	12.6	10	38.8
ePt-638325	8	35.4	9	61.8
ePt-638853	3	88.3	7	51.2
ePt-638863	4	3.2	8	31.1
ePt-639810	10	13.7	11	86.7
ePt-639925	10	10.5	11	86.6
ePt-640753	3	29.1	9	26.1
ePt-640754	1	26.2	4	58.2
ePt-640819	3	55.5	11	58.7
ePt-641396	2	79.2	11	46.8
ePt-641492	9	49.4	10	32.7
ePt-641571	7	65.3	8	74.8
ePt-641578	6	1.8	8	88.9
ePt-641639	4	48.0	8	55.1
ePt-641685	7	65.3	8	74.3
ePt-642931	5	39.3	6	15.0
ePt-643056	9	49.4	10	32.7
ePt-643259	3	58.2	5	12.4

ePt-643271	3	57.4	5	1.5				
ePt-643707	3	56.5	5	1.6				
ePt-643832	5	44.5	8	90.1				
ePt-643890	3	53.8	8	106.0				
ePt-643897	6	37.7	8	90.1				
Markers mapped to three linkage groups								
ePt-574238	2	16.6	5	85.6	6	26.8		
ePt-643458	6	86.6	7	0.0	8	85.0		
Markers mapped to four linkage groups								
ePt-503174	2	15.5	3	41.5	5	72.2	8	88.2
ePt-568818	2	55.1	3	85.5	5	77.7	7	80.5
ePt-637610	2	60.2	3	40.9	5	72.1	8	90.1
ePt-637861	2	60.7	3	40.9	5	72.1	8	90.1

Table S4.1 Sample name, locality and STRUCTURE population code information of 444 eucalypt samples.

Species and sample name[a]	STRUCTURE analysis[b]		Locality	State	Population / subspecies[c]
	Seq	Pop #			
E. globulus					
GL_2628	1	1	Recherche Bay	TAS	Recherche Bay
GL_2796	2	1	Recherche Bay	TAS	Recherche Bay
GL_1531	3	1	Dover	TAS	southern Tasmania
GL_1537	4	1	Dover	TAS	southern Tasmania
GL_1521	5	1	Geeveston	TAS	southern Tasmania
GL_1529	6	1	Geeveston	TAS	southern Tasmania
GL_1538	7	1	South Bruny Island	TAS	southern Tasmania
GL_1544	8	1	South Bruny Island	TAS	southern Tasmania
GL_1348	9	1	Tinderbox	TAS	southern Tasmania
GL_1805	10	1	Hobart (Chimney Pot Hill)	TAS	southeast Tasmania
GL_1811	11	1	Hobart (Chimney Pot Hill)	TAS	southeast Tasmania
GL_2913	12	1	Mt Dromedary	TAS	Dromedary
GL_2932	13	1	Platform Peak	TAS	Dromedary
GL_1760	14	2	Jericho	TAS	northeast Tasmania
GL_1764	15	2	Jericho	TAS	northeast Tasmania
GL_1003	16	2	White Beach	TAS	southern Tasmania
GL_1046	17	2	Murdunna	TAS	southern Tasmania
GL_1056	18	2	Break-me-neck-Hill	TAS	southeast Tasmania
GL_1193	19	2	South Buckland (military)	TAS	southeast Tasmania
GL_1796	20	2	Maria Island south	TAS	southeast Tasmania
GL_1769	21	2	Maria Island north	TAS	southeast Tasmania
GL_1754	22	2	Triabunna	TAS	southeast Tasmania
GL_1756	23	2	Triabunna	TAS	southeast Tasmania
GL_1745	24	3	Mayfield south	TAS	northeast Tasmania
GL_1750	25	3	Mayfield south	TAS	northeast Tasmania
GL_1736	26	3	Mayfield north	TAS	northeast Tasmania
GL_1740	27	3	Mayfield north	TAS	northeast Tasmania
GL_1655	28	3	Cape Tourville (tall)	TAS	northeast Tasmania
GL_1345	29	3	Cape Tourville (dwarf)	TAS	northeast Tasmania
GL_1719	30	3	Pepper Hill	TAS	northeast Tasmania
GL_1722	31	3	Pepper Hill	TAS	northeast Tasmania
GL_1726	32	3	Humbug Hill	TAS	northeast Tasmania
GL_1731	33	3	Humbug Hill	TAS	northeast Tasmania
GL_2071	34	4	Macquarie Harbour	TAS	western Tasmania
GL_2079	35	4	Macquarie Harbour	TAS	western Tasmania
GL_3403	36	4	Port Davey	TAS	western Tasmania
GL_3422	37	4	Port Davey	TAS	western Tasmania
GL_2052	38	4	Badgers Creek	TAS	western Tasmania
GL_2057	39	4	Badgers Creek	TAS	western Tasmania
GL_2058	40	4	Badgers Creek	TAS	western Tasmania
GL_2059	41	4	Little Henty River	TAS	western Tasmania
GL_2064	42	4	Little Henty River	TAS	western Tasmania
GL_1816	43	5	South King Island	TAS	King Island
GL_1821	44	5	South King Island	TAS	King Island
GL_2124	45	5	Central King Island	TAS	King Island
GL_2132	46	5	Central King Island	TAS	King Island
GL_2136	47	5	Central north King Island	TAS	King Island
GL_2137	48	5	Central north King Island	TAS	King Island
GL_2141	49	5	North King Island	TAS	King Island

	GL_2145	50	5	North King Island	TAS	King Island
	GL_1406	51	6	Cape Barren Island	TAS	Furneaux
	GL_1432	52	6	Cape Barren Island	TAS	Furneaux
	GL_1332	53	6	Flinders Island south	TAS	Furneaux
	GL_1334	54	6	Flinders Island south	TAS	Furneaux
	GL_1390	55	6	Flinders Island central	TAS	Furneaux
	GL_1403	56	6	Flinders Island central	TAS	Furneaux
	GL_1391	57	6	Flinders Island north	TAS	Furneaux
	GL_1392	58	6	Flinders Island north	TAS	Furneaux
	GL_2577	59	7	Wilsons Promontory (lighthouse)	VIC	Wilsons Promontory
	GL_2605	60	7	Wilsons Promontory (lighthouse)	VIC	Wilsons Promontory
	GL_9904	61	7	Tidal River	VIC	Wilsons Promontory
	GL_9917	62	7	Tidal River	VIC	Wilsons Promontory
	GL_9923	63	8	Phillip Island	VIC	southern Gippsland
	GL_9940	64	8	Phillip Island	VIC	southern Gippsland
	GL_1254	65	8	Fish Creek	VIC	southern Gippsland
	GL_1258	66	8	Fish Creek	VIC	southern Gippsland
	GL_1261	67	8	Toora	VIC	southern Gippsland
	GL_1267	68	8	Toora	VIC	southern Gippsland
	GL_1268	69	8	AlbWest /Welshpool/Hedley	VIC	southern Gippsland
	GL_1273	70	8	AlbWest/Welshpool/Hedley	VIC	southern Gippsland
	GL_1139	71	9	Jeeralang	VIC	Strzelecki Ranges
	GL_1160	72	9	Jeeralang	VIC	Strzelecki Ranges
	GL_1554	73	11	Lorne	VIC	eastern Otways
	GL_1558	74	11	Lorne	VIC	eastern Otways
	GL_1572	75	11	Jamieson Creek	VIC	eastern Otways
	GL_1580	76	11	Cape Patton	VIC	eastern Otways
	GL_1585	77	11	Cape Patton	VIC	eastern Otways
	GL_1588	78	10	Cannan Spur	VIC	western Otways
	GL_1593	79	10	Cannan Spur	VIC	western Otways
	GL_1594	80	10	Parker Spur	VIC	western Otways
	GL_1599	81	10	Parker Spur	VIC	western Otways
	GL_1704	82	10	Otway State Forest	VIC	western Otways
	GL_1708	83	10	Otway State Forest	VIC	western Otways
	GL_1712	84	10	Lavers Hill	VIC	western Otways
E. nitens						
	N_TO110	85	12	Spion Kopje	VIC	Central Victoria - south
	N_TO108	86	12	Spion Kopje	VIC	Central Victoria - south
	N_TO102	87	12	West Kopje	VIC	Central Victoria - south
	N_TO100	88	12	West Kopje	VIC	Central Victoria - south
	N_TO124	89	12	Starling Gap	VIC	Central Victoria - south
	N_TO118	90	12	Starling Gap	VIC	Central Victoria - south
	N_2087676	91	12	Mount Erica	VIC	Central Victoria - south
	N_2087580	92	12	Mount Erica	VIC	Central Victoria - south
	N_2087607	93	12	Mt St Gwinear Road	VIC	Central Victoria - south
	N_2087532	94	12	Mt St Gwinear Road	VIC	Central Victoria - south
	N_2087530	95	12	Newlands Road	VIC	Central Victoria - south
	N_TV53	96	12	Upper Thomson Valley	VIC	Central Victoria - south
	N_2087596	97	12	Thomson Valley - Creeks	VIC	Central Victoria - south
	N_2087518	98	12	Thomson Valley - Creeks	VIC	Central Victoria - south
	N_TO11	99	12	Mount Horsfall	VIC	Central Victoria - south
	N_TO1	100	12	Mount Horsfall	VIC	Central Victoria - south
	N_2087634	101	12	Marshall Spur	VIC	Central Victoria - south
	N_2087514	102	12	Number 3 Road	VIC	Central Victoria - south

N_2087622	103	12	Toorongo Town	VIC	Central Victoria - south
N_H23	104	12	Toorongo north	VIC	Central Victoria - south
N_H14	105	12	Toorongo north	VIC	Central Victoria - south
N_2087716	106	12	Toorongo north	VIC	Central Victoria - south
N_2087570	107	12	Toorongo south	VIC	Central Victoria - south
N_TO135	108	12	Ben Cairn	VIC	Central Victoria - south
N_TO128	109	12	Ben Cairn	VIC	Central Victoria - south
N_TO127	110	12	Ben Cairn	VIC	Central Victoria - south
N_TO140	111	12	Bonna Buang	VIC	Central Victoria - south
N_TO138	112	12	Bonna Buang	VIC	Central Victoria - south
N_TO137	113	12	Bonna Buang	VIC	Central Victoria - south
N_RU103	114	12	Monda Road	VIC	Central Victoria - south
N_RU102	115	12	Monda Road	VIC	Central Victoria - south
N_Ru114	116	12	Mount Tanglefoot	VIC	Central Victoria - south
N_RU108	117	12	Mount Tanglefoot	VIC	Central Victoria - south
N_RU101	118	12	Mount Tanglefoot	VIC	Central Victoria - south
N_2087758	119	13	Royston river	VIC	Central Victoria - north
N_RU78	120	13	Quartz Creek	VIC	Central Victoria - north
N_2087603	121	13	Quartz Creek	VIC	Central Victoria - north
N_2087543	122	13	Snobs Creek	VIC	Central Victoria - north
N_2087697	123	13	Little River	VIC	Central Victoria - north
N_2087541	124	13	Little River	VIC	Central Victoria - north
N_2087630	125	13	Barnewall Plains	VIC	Central Victoria - north
N_2087528	126	13	Barnewall Plains	VIC	Central Victoria - north
N_TO67	127	13	Mount Gregory	VIC	Central Victoria - north
N_TO64	128	13	Mount Gregory	VIC	Central Victoria - north
N_TO60	129	13	Mount Gregory	VIC	Central Victoria - north
N_2087610	130	13	Mount Useful – southeast	VIC	Central Victoria - north
N_2087562	131	13	Mount Useful – southeast	VIC	Central Victoria - north
N_2087567	132	13	Mount Useful – south	VIC	Central Victoria - north
N_MA104	133	13	Mount Useful – east	VIC	Central Victoria - north
N_MA101	134	13	Mount Useful – east	VIC	Central Victoria - north
N_MA80	135	13	Mount Useful – north	VIC	Central Victoria - north
N_2087694	136	13	Mt Skene - Lazarini Creek	VIC	Central Victoria - north
N_2087664	137	13	Mt Skene - Lazarini Creek	VIC	Central Victoria - north
N_2087521	138	13	Mt Skene - Lazarini Creek	VIC	Central Victoria - north
N_MA49	139	13	Mt Skene / Barkly River - E	VIC	Central Victoria - north
N_MA48	140	13	Mt Skene / Barkly River - E	VIC	Central Victoria - north
N_MA2	141	14	Connors Plains – southeast	VIC	Connors Plains
N_2087536	142	14	Connors Plains – southeast	VIC	Connors Plains
N_MA15	143	14	Connors Plains - plateau	VIC	Connors Plains
N_2087721	144	14	Connors Plains - plateau	VIC	Connors Plains
N_2087638	145	14	Connors Plains – northwest	VIC	Connors Plains
N_2087544	146	15	Mt Wellington	VIC	Mt Wellington
N_2087526	147	16	Victoria – unknown	VIC	Central Victoria -
N_2087525	148	16	Victoria – unknown	VIC	Central Victoria -
N_2087524	149	16	Victoria – unknown	VIC	Central Victoria -
N_2358996	150	17	Nimitabel	NSW	southern NSW
N_2358669	151	17	Nimitabel	NSW	southern NSW
N_2087749	152	17	Nimitabel	NSW	southern NSW
N_2087553	153	17	Nimitabel	NSW	southern NSW
N_2087551	154	17	Nimitabel	NSW	southern NSW
N_2087635	155	17	Unknown	NSW	southern NSW
N_2359047	156	17	Tallaganda	NSW	southern NSW
N_2358790	157	17	Tallaganda	NSW	southern NSW
N_2358691	158	17	Tallaganda	NSW	southern NSW

N_2358644	159	17	Tallaganda	NSW	southern NSW
N_2358402	160	17	Tallaganda	NSW	southern NSW
N_2358698	161	18	Barrington Tops	NSW	northern NSW
N_2358693	162	18	Barrington Tops	NSW	northern NSW
N_2358626	163	18	Barrington Tops	NSW	northern NSW
N_2358490	164	18	Barrington Tops	NSW	northern NSW
N_2359035	165	18	Ebor	NSW	northern NSW
N_2358918	166	18	Ebor	NSW	northern NSW
N_2358852	167	18	Ebor	NSW	northern NSW
N_2358849	168	18	Ebor	NSW	northern NSW
N_2358346	169	18	Ebor	NSW	northern NSW
E. camaldulensis					
C_74_425C	170	19	Laura River	Qld	*simulata*
C_76_400	171	19	Normanby River	Qld	*simulata*
C_3_450	172	20	Gilbert River	Qld	*acuta* (northern)
C_6_415	173	20	Morehead River	Qld	*acuta* (northern)
C_7_390	174	20	Petford	Qld	*acuta* (northern)
C_7_391	175	20	Petford	Qld	*acuta* (northern)
C_1_MM3114	176	21	Adavale	Qld	*acuta* (southern)
C_2_3189	177	21	Angellala River	Qld	*acuta* (southern)
C_5_3166	178	21	Moonie River	Qld	*acuta* (southern)
C_5_3170	179	21	Moonie River	Qld	*acuta* (southern)
C_15_3156	180	22	Dirrandanbi	Qld	*camaldulensis*
C_10_3340	181	22	Aberdeen	NSW	*camaldulensis*
C_28_CS168	182	22	Wellington	NSW	*camaldulensis*
C_27_163	183	22	Warren	NSW	*camaldulensis*
C_26_3152	184	22	Warraweena	NSW	*camaldulensis*
C_13_155	185	22	Condobolin	NSW	*camaldulensis*
C_30_125	186	22	Wilcannia	NSW	*camaldulensis*
C_17_SC134	187	22	Hillston	NSW	*camaldulensis*
C_21_114	188	22	Menindee	NSW	*camaldulensis*
C_29_86	189	22	Wentworth	NSW	*camaldulensis*
C_25_CS58	190	22	Towong	NSW	*camaldulensis*
C_11_77	191	22	Barmah	NSW	*camaldulensis*
C_31_271	192	22	Wirrengren Plain	Vic	*camaldulensis*
C_20_PM295	193	22	Lake Albacutya	Vic	*camaldulensis*
C_16_334	194	22	Elmhurst	Vic	*camaldulensis*
C_18_323	195	22	Horsham	Vic	*camaldulensis*
C_23_3026	196	22	Morgan	SA	*camaldulensis*
C_24_3012	197	22	Mypolonga	SA	*camaldulensis*
C_24_3016	198	22	Mypolonga	SA	*camaldulensis*
C_22_13253-1	199	22	Minlaton	SA	*camaldulensis*
C_19_KI2	200	22	Kangaroo Island	SA	*camaldulensis*
C_12_3005	201	22	Coffin Bay	SA	*camaldulensis*
C_32_2974	202	23	Boolcunda Creek	SA	*minima*
C_33_2994	203	23	Bunyeroo Creek	SA	*minima*
C_34_2966	204	23	Horrocks Pass	SA	*minima*
C_35_2958	205	23	Huddleston	SA	*minima*
C_35_2959	206	23	Huddleston	SA	*minima*
C_82_2981	207	24	Emu Creek	SA	*arida*
C_79_107	208	24	Broken River	NSW	*arida*
C_88_SC92	209	24	Silverton	NSW	*arida*
C_80_MM320	210	24	Bulloo River	Qld	*arida*
C_91_3197	211	24	Ward River	Qld	*arida*
C_93_3045	212	24	Windorah	Qld	*arida*
C_78_3109	213	24	Boulia	Qld	*arida*

C_85_3066	214	24	Muttaburra	Qld	*arida*
C_77_MM322	215	24	Arthur Creek	Qld	*arida*
C_84_460B	216	24	Lander River	NT	*arida*
C_81_3327	217	24	Coongra Creek	SA	*arida*
C_86_3253	218	24	Palmer River	NT	*arida*
C_83_3267	219	24	Giles Creek	WA	*arida*
C_90_3314	220	24	Warbuton	WA	*arida*
C_89_2780	221	24	Station Creek	WA	*arida*
C_87_470C	222	24	Rudall River	WA	*arida*
C_92_2946	223	24	Wiluna	WA	*arida*
C_72_2823	224	25	Newman	WA	*refulgens*
C_73_2809	225	25	Nullagine	WA	*refulgens*
C_65_MM286	226	25	Gorge Junction	WA	*refulgens*
C_70_MM287	227	25	Moorarie	WA	*refulgens*
C_67_2929	228	25	Meeberrie	WA	*refulgens*
C_71_2798	229	25	Muccan	WA	*refulgens*
C_62_2792	230	25	De Grey River	WA	*refulgens*
C_68_2835	231	25	Millstream	WA	*refulgens*
C_66_2858	232	25	Kooline	WA	*refulgens*
C_63_2888	233	25	Elong	WA	*refulgens*
C_60_2900	234	25	Bidgemia	WA	*refulgens*
C_61_2910	235	25	Carnarvon	WA	*refulgens*
C_69_2850	236	25	Mindaroo	WA	*refulgens*
C_64_2838	237	25	Fortescue	WA	*refulgens*
C_40_2897	238	26	Fitzroy River lower	WA	*obtusa*
C_39_2880	239	26	Fitzroy Crossing	WA	*obtusa*
C_42_2905	240	26	Gibb River	WA	*obtusa*
C_46_2873C	241	26	Margaret River	WA	*obtusa*
C_50_2861B	242	26	Ord River upper	WA	*obtusa*
C_49_2909B	243	26	Ord River lower	WA	*obtusa*
C_58_2946	244	26	Victoria River upper	NT	*obtusa*
C_57_2920	245	26	Victoria River lower	NT	*obtusa*
C_56_2937B	246	26	Victoria River Downs	NT	*obtusa*
C_38_ER9	247	26	Edith River	NT	*obtusa*
C_47_MR6	248	26	Mary River	NT	*obtusa*
C_51_UR2	249	26	Roper River	NT	*obtusa*
C_53_SgC10	250	26	Strangeways Creek	NT	*obtusa*
C_37_MM323	251	26	Daly Waters Creek	NT	*obtusa*
C_44_HR6	252	26	Hodgson River	NT	*obtusa*
C_54_SC2	253	26	Surprise Creek	NT	*obtusa*
C_55_TC3B	254	26	Tennant Creek	NT	*obtusa*
C_52_SC2	255	26	Settlement Creek	NT	*obtusa*
C_59_3094	256	26	Whistler Creek	Qld	*obtusa*
C_41_2764	257	26	Floraville	Qld	*obtusa*
C_45_2754	258	26	Kamilaroi	Qld	*obtusa*
C_48_2772	259	26	Mt Isa	Qld	*obtusa*
C_43_3082	260	26	Glen Gorge	Qld	*obtusa*
C_36_3060	261	26	Baroota Waterhole	Qld	*obtusa*
E. grandis					
GR_DP192	262	27	Bulahdelah	NSW	southern NSW
GR_DP109	263	27	Bulahdelah	NSW	southern NSW
GR_DP108	264	27	Bulahdelah	NSW	southern NSW
GR_DP116	265	27	Wang Wauk SF	NSW	southern NSW
GR_DP191	266	27	Mt. George Taree	NSW	southern NSW
GR_DP190	267	27	Mt. George Taree	NSW	southern NSW
GR_DP189	268	27	Mt. George Taree	NSW	southern NSW

GR_DP188	269	27	Mt. George Taree	NSW	southern NSW
GR_DP186	270	27	Mt. George Taree	NSW	southern NSW
GR_DP185	271	27	Mt. George Taree	NSW	southern NSW
GR_DP183	272	27	Mt. George Taree	NSW	southern NSW
GR_DP182	273	27	Mt. George Taree	NSW	southern NSW
GR_DP181	274	27	Mt. George Taree	NSW	southern NSW
GR_DP110	275	27	Mt. George Taree	NSW	southern NSW
GR_DP105	276	27	Mt. George Taree	NSW	southern NSW
GR_DP180	277	27	Taree	NSW	southern NSW
GR_DP102	278	27	Taree	NSW	southern NSW
GR_DP179	279	27	Lake Cathie	NSW	southern NSW
GR_DP106	280	27	Lake Cathie	NSW	southern NSW
GR_DP103	281	27	Lake Cathie	NSW	southern NSW
GR_DP187	282	27	Wauchope	NSW	southern NSW
GR_DP184	283	27	Wauchope	NSW	southern NSW
GR_DP111	284	27	Wauchope	NSW	southern NSW
GR_DP107	285	27	Wauchope	NSW	southern NSW
GR_DP104	286	27	Wauchope	NSW	southern NSW
GR_DP101	287	27	Wauchope	NSW	southern NSW
GR_DP172	288	28	Urunga	NSW	Coffs Harbour
GR_DP171	289	28	Urunga	NSW	Coffs Harbour
GR_DP168	290	28	Urunga	NSW	Coffs Harbour
GR_DP159	291	28	Urunga	NSW	Coffs Harbour
GR_DP173	292	28	Pine Creek SF	NSW	Coffs Harbour
GR_DP170	293	28	Pine Creek SF	NSW	Coffs Harbour
GR_DP163	294	28	Pine Creek SF	NSW	Coffs Harbour
GR_DP176	295	28	Boambie Coffs Harbour	NSW	Coffs Harbour
GR_DP165	296	28	Oraba Coffs Harbour	NSW	Coffs Harbour
GR_DP164	297	28	Oraba Coffs Harbour	NSW	Coffs Harbour
GR_DP167	298	28	Coffs Harbour	NSW	Coffs Harbour
GR_DP166	299	28	Coffs Harbour	NSW	Coffs Harbour
GR_DP161	300	28	Coffs Harbour	NSW	Coffs Harbour
GR_DP160	301	28	Coffs Harbour	NSW	Coffs Harbour
GR_DP174	302	28	Lower Bucca	NSW	Coffs Harbour
GR_DP162	303	28	Lower Bucca	NSW	Coffs Harbour
GR_DP175	304	28	Conglomerate SF	NSW	Coffs Harbour
GR_DP178	305	28	Toonumba	NSW	Coffs Harbour
GR_DP177	306	28	Yabra/Urbenville	NSW	Coffs Harbour
GR_DP155	307	29	Belthorpe	Qld	Gympie
GR_DP154	308	29	Belthorpe	Qld	Gympie
GR_DP151	309	29	Belthorpe	Qld	Gympie
GR_DP146	310	29	Belthorpe	Qld	Gympie
GR_DP144	311	29	Belthorpe	Qld	Gympie
GR_DP141	312	29	Belthorpe	Qld	Gympie
GR_DP114	313	29	Belthorpe	Qld	Gympie
GR_DP158	314	29	Kenilworth	Qld	Gympie
GR_DP157	315	29	Kenilworth	Qld	Gympie
GR_DP156	316	29	Kenilworth	Qld	Gympie
GR_DP153	317	29	Kenilworth	Qld	Gympie
GR_DP152	318	29	Kenilworth	Qld	Gympie
GR_DP148	319	29	Kenilworth	Qld	Gympie
GR_DP145	320	29	Kenilworth	Qld	Gympie
GR_DP117	321	29	Kenilworth	Qld	Gympie
GR_DP150	322	29	Woondum/Gympie	Qld	Gympie
GR_DP143	323	29	Woondum/Gympie	Qld	Gympie
GR_DP142	324	29	Woondum/Gympie	Qld	Gympie

GR_DP115	325	29	Woondum/Gympie	Qld	Gympie
GR_DP147	326	29	Veteran Gympie	Qld	Gympie
GR_DP134	327	30	Mt Windsor	Qld	Atherton
GR_DP138	328	30	Townsville	Qld	Atherton
GR_DP137	329	30	Townsville	Qld	Atherton
GR_DP133	330	30	Townsville	Qld	Atherton
GR_DP124	331	30	Townsville	Qld	Atherton
GR_DP126	332	30	Kennedy	Qld	Atherton
GR_DP131	333	30	Wandecla	Qld	Atherton
GR_DP125	334	30	Wandecla	Qld	Atherton
GR_DP122	335	30	Wandecla	Qld	Atherton
GR_DP121	336	30	Wandecla	Qld	Atherton
GR_DP118	337	30	Wandecla	Qld	Atherton
GR_DP132	338	30	Ravenshoe	Qld	Atherton
GR_DP129	339	30	Ravenshoe	Qld	Atherton
GR_DP136	340	30	Baldy SF	Qld	Atherton
GR_DP130	341	30	Baldy SF	Qld	Atherton
GR_DP120	342	30	Baldy SF	Qld	Atherton
GR_DP113	343	30	Baldy SF	Qld	Atherton
GR_DP112	344	30	Baldy SF	Qld	Atherton
GR_DP128	345	30	Atherton	Qld	Atherton
GR_DP127	346	30	Atherton	Qld	Atherton
GR_DP119	347	30	Atherton	Qld	Atherton
GR_DP140	348	30	Mareeba	Qld	Atherton
GR_DP139	349	30	Mareeba	Qld	Atherton
GR_DP135	350	30	Mareeba	Qld	Atherton
GR_DP123	351	30	Mareeba	Qld	Atherton
E. urophylla					
U_DP027	352	31	Bonleu	N/A	Timor
U_DP028	353	31	Bonleu	N/A	Timor
U_DP020	354	31	Lelobatan	N/A	Timor
U_DP021	355	31	Lelobatan	N/A	Timor
U_DP094	356	31	Leloboko	N/A	Timor
U_DP091	357	31	Mollo	N/A	Timor
U_DP092	358	31	Mollo	N/A	Timor
U_DP022	359	31	Nuafin	N/A	Timor
U_DP023	360	31	Nuafin	N/A	Timor
U_DP024	361	31	Nuafin	N/A	Timor
U_DP096	362	31	Fatumnase	N/A	Timor
U_DP016	363	31	Naususu	N/A	Timor
U_DP017	364	31	Naususu	N/A	Timor
U_DP095	365	31	A. Esrael	N/A	Timor
U_DP093	366	31	Lelobatang	N/A	Timor
U_DP025	367	31	Tutem	N/A	Timor
U_DP026	368	31	Tutem	N/A	Timor
U_DP018	369	31	Tune	N/A	Timor
U_DP019	370	31	Tune	N/A	Timor
U_DP089	371	32	Lere-Baukrenget	N/A	Flores
U_DP090	372	32	Lere-Baukrenget	N/A	Flores
U_DP054	373	32	Natakoli	N/A	Flores
U_DP055	374	32	Natakoli	N/A	Flores
U_DP044	375	32	Ile Nggele	N/A	Flores
U_DP045	376	32	Ile Nggele	N/A	Flores
U_DP046	377	32	Ile Nggele	N/A	Flores
U_DP047	378	32	Kilawair	N/A	Flores
U_DP048	379	32	Kilawair	N/A	Flores

U_DP052	380	32	Kolibuluk	N/A	Flores
U_DP053	381	32	Kolibuluk	N/A	Flores
U_DP050	382	32	Hokeng	N/A	Flores
U_DP051	383	32	Hokeng	N/A	Flores
U_DP014	384	33	Kawela	N/A	Adonara
U_DP015	385	33	Kawela	N/A	Adonara
U_DP012	386	33	Lamahela	N/A	Adonara
U_DP013	387	33	Lamahela	N/A	Adonara
U_DP008	388	33	Watololong	N/A	Adonara
U_DP009	389	33	Watololong	N/A	Adonara
U_DP010	390	33	Gonehama	N/A	Adonara
U_DP011	391	33	Gonehama	N/A	Adonara
U_DP003	392	33	Muda	N/A	Adonara
U_DP004	393	33	Muda	N/A	Adonara
U_DP001	394	33	Doken	N/A	Adonara
U_DP002	395	33	Doken	N/A	Adonara
U_DP006	396	33	Lamalota	N/A	Adonara
U_DP007	397	33	Lamalota	N/A	Adonara
U_DP066	398	34	Puor	N/A	Lomblen
U_DP067	399	34	Puor	N/A	Lomblen
U_DP058	400	34	Jontona	N/A	Lomblen
U_DP059	401	34	Jontona	N/A	Lomblen
U_DP060	402	34	Jontona	N/A	Lomblen
U_DP056	403	34	Padekluwa	N/A	Lomblen
U_DP057	404	34	Padekluwa	N/A	Lomblen
U_DP064	405	34	Ile Kerbau	N/A	Lomblen
U_DP065	406	34	Ile Kerbau	N/A	Lomblen
U_DP068	407	34	Ile Ape	N/A	Lomblen
U_DP069	408	34	Ile Ape	N/A	Lomblen
U_DP061	409	34	Labalekan	N/A	Lomblen
U_DP062	410	34	Labalekan	N/A	Lomblen
U_DP063	411	34	Labalekan	N/A	Lomblen
U_DP079	412	35	Mauta	N/A	Pantar
U_DP080	413	35	Mauta	N/A	Pantar
U_DP081	414	35	Mauta	N/A	Pantar
U_DP085	415	35	Delaki	N/A	Pantar
U_DP086	416	35	Delaki	N/A	Pantar
U_DP087	417	35	Beangonong	N/A	Pantar
U_DP088	418	35	Beangonong	N/A	Pantar
U_DP082	419	35	Lalapang	N/A	Pantar
U_DP083	420	35	Lalapang	N/A	Pantar
U_DP084	421	35	Lalapang	N/A	Pantar
U_DP078	422	36	Watakika	N/A	Alor
U_DP075	423	36	Pintu Mas	N/A	Alor
U_DP076	424	36	Pintu Mas	N/A	Alor
U_DP077	425	36	Pintu Mas	N/A	Alor
U_DP070	426	36	Mainang	N/A	Alor
U_DP071	427	36	Mainang	N/A	Alor
U_DP072	428	36	Apui	N/A	Alor
U_DP073	429	36	Apui	N/A	Alor
U_DP074	430	36	Apui	N/A	Alor
U_DP040	431	37	Nesunhuhun	N/A	Wetar
U_DP041	432	37	Nesunhuhun	N/A	Wetar
U_DP037	433	37	Elun Kripas	N/A	Wetar
U_DP038	434	37	Elun Kripas	N/A	Wetar
U_DP042	435	37	Nakana Ulam	N/A	Wetar

U_DP043	436	37	Nakana Ulam	N/A	Wetar
U_DP035	437	37	Alasannaru	N/A	Wetar
U_DP036	438	37	Alasannaru	N/A	Wetar
U_DP029	439	37	Puaanan	N/A	Wetar
U_DP030	440	37	Puaanan	N/A	Wetar
U_DP031	441	37	Remamea	N/A	Wetar
U_DP032	442	37	Remamea	N/A	Wetar
U_DP033	443	37	Talianan	N/A	Wetar
U_DP034	444	37	Talianan	N/A	Wetar

[a]Sample name. Samples have been given a species prefix (e.g. GL for *E. globulus*). *Eucalyptus camaldulensis* sample names have been coded as; species prefix followed by population code from Butcher *et al* (2009) and CSIRO sample number. [b]Sample population code (Pop #) and sequence (Seq) number in 444 sample x 2207 marker STRUCTURE analysis. [c]Population follows Table 4.1. SF=state forest.

Supplementary figures

Following pages:

Figure S1.1A-D Consensus linkage map constructed in the *E. globulus* Lighthouse F_2 family. The map contains 1060 markers; including 1010 DArT markers ('ePt' prefix) and 50 microsatellite markers. Linkage group numbering and orientation follows Brondani *et al.* (2002; 2006) and map units are in Kosambi centiMorgan. Markers added to the 753-marker framework map are underlined. Loci with distorted segregation ratios are indicated by asterisk; *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ****$P \leq 0.0001$.

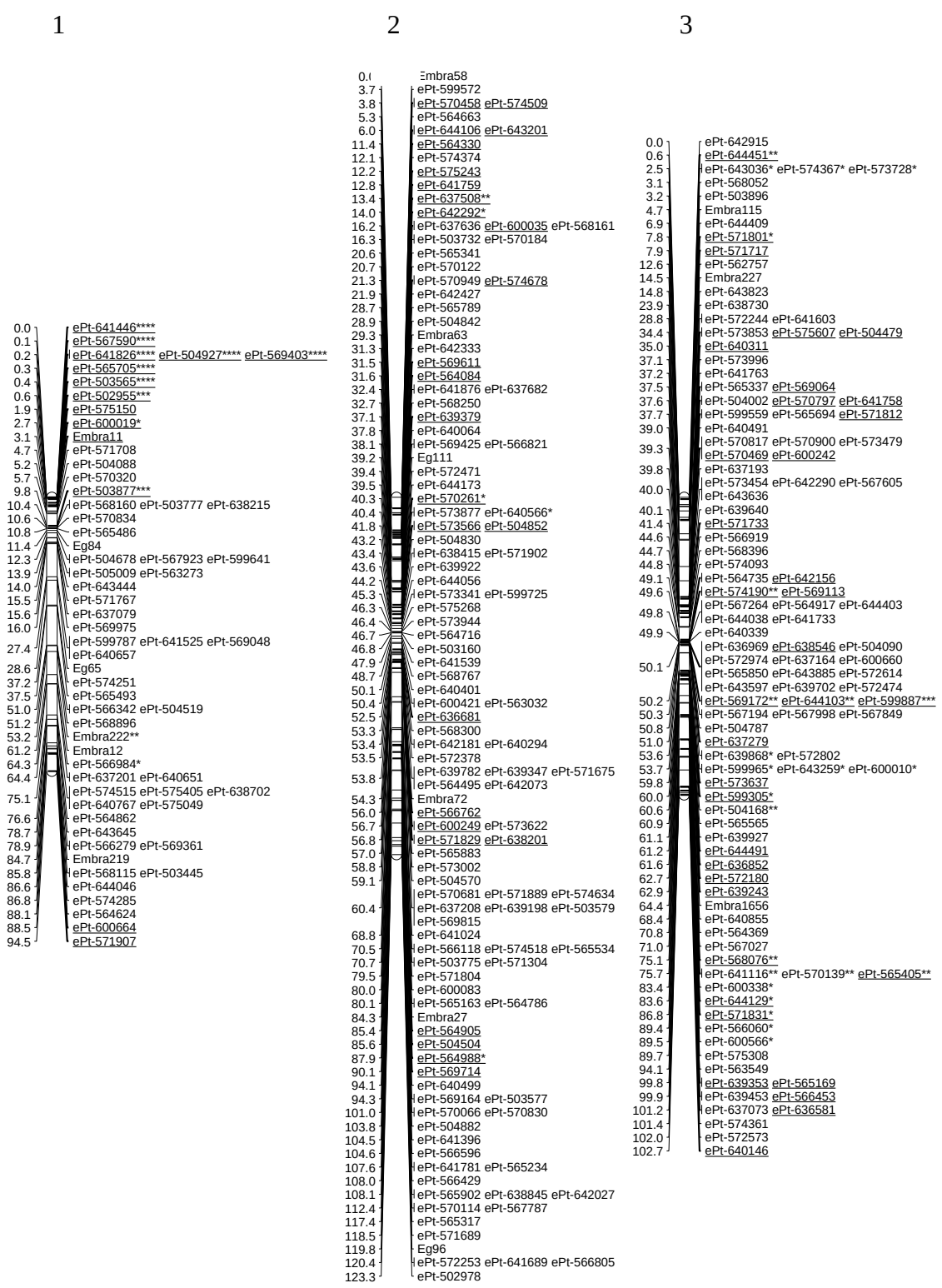
1
2
3
0.0 ePt-641446****
0.1 ePt-567590****
0.2 ePt-641826**** ePt-504927**** ePt-569403****
0.3 ePt-565705****
0.4 ePt-503565****
0.6 ePt-502955***
1.9 ePt-575150
2.7 ePt-600019*
3.1 Embra11
4.7 ePt-571708
5.2 ePt-504088
5.7 ePt-570320
9.8 ePt-503877***
10.4 ePt-568160 ePt-503777 ePt-638215
10.6 ePt-570834
10.8 ePt-565486
11.4 Eg84
12.3 ePt-504678 ePt-567923 ePt-599641
13.9 ePt-505009 ePt-563273
14.0 ePt-643444
15.5 ePt-571767
15.6 ePt-637079
16.0 ePt-569975
27.4 ePt-599787 ePt-641525 ePt-569048
ePt-640657
28.6 Eg65
37.2 ePt-574251
37.5 ePt-565493
51.0 ePt-566342 ePt-504519
51.2 ePt-568896
53.2 Embra222**
61.2 Embra12
64.3 ePt-566984*
64.4 ePt-637201 ePt-640651
75.1 ePt-574515 ePt-575405 ePt-638702
ePt-640767 ePt-575049
76.6 ePt-564862
78.7 ePt-643645
78.9 ePt-566279 ePt-569361
84.7 Embra219
85.8 ePt-568115 ePt-503445
86.6 ePt-644046
86.8 ePt-574285
88.1 ePt-564624
88.5 ePt-600664
94.5 ePt-571907
0.0 Embra58
3.7 ePt-599572
3.8 ePt-570458 ePt-574509
5.3 ePt-564663
6.0 ePt-644106 ePt-643201
11.4 ePt-564330
12.1 ePt-574374
12.2 ePt-575243
12.8 ePt-641759
13.4 ePt-637508**
14.0 ePt-642292*
16.2 ePt-637636 ePt-600035 ePt-568161
16.3 ePt-503732 ePt-570184
20.6 ePt-565341
20.7 ePt-570122
21.3 ePt-570949 ePt-574678
21.9 ePt-642427
28.7 ePt-565789
28.9 ePt-504842
29.3 Embra63
31.3 ePt-642333
31.5 ePt-569611
31.6 ePt-564084
32.4 ePt-641876 ePt-637682
32.7 ePt-568250
37.1 ePt-639379
37.8 ePt-640064
38.1 ePt-569425 ePt-566821
39.2 Eg111
39.4 ePt-572471
39.5 ePt-644173
40.3 ePt-570261*
40.4 ePt-573877 ePt-640566*
41.8 ePt-573566 ePt-504852
43.2 ePt-504830
43.4 ePt-638415 ePt-571902
43.6 ePt-639922
44.2 ePt-644056
45.3 ePt-573341 ePt-599725
46.3 ePt-575268
46.4 ePt-573944
46.7 ePt-564716
46.8 ePt-503160
47.9 ePt-641539
48.7 ePt-568767
50.1 ePt-640401
50.4 ePt-600421 ePt-563032
52.5 ePt-636681
53.3 ePt-568300
53.4 ePt-642181 ePt-640294
53.5 ePt-572378
53.8 ePt-639782 ePt-639347 ePt-571675
ePt-564495 ePt-642073
54.3 Embra72
56.0 ePt-566762
56.7 ePt-600249 ePt-573622
56.8 ePt-571829 ePt-638201
57.0 ePt-565883
58.8 ePt-573002
59.1 ePt-504570
60.4 ePt-570681 ePt-571889 ePt-574634
ePt-637208 ePt-639198 ePt-503579
ePt-569815
68.8 ePt-641024
70.5 ePt-566118 ePt-574518 ePt-565534
70.7 ePt-503775 ePt-571304
79.5 ePt-571804
80.0 ePt-600083
80.1 ePt-565163 ePt-564786
84.3 Embra27
85.4 ePt-564905
85.6 ePt-504504
87.9 ePt-564988*
90.1 ePt-569714
94.1 ePt-640499
94.3 ePt-569164 ePt-503577
101.0 ePt-570066 ePt-570830
103.8 ePt-504882
104.5 ePt-641396
104.6 ePt-566596
107.6 ePt-641781 ePt-565234
108.0 ePt-566429
108.1 ePt-565902 ePt-638845 ePt-642027
112.4 ePt-570114 ePt-567787
117.4 ePt-565317
118.5 ePt-571689
119.8 Eg96
120.4 ePt-572253 ePt-641689 ePt-566805
123.3 ePt-502978
0.0 ePt-642915
0.6 ePt-644451**
2.5 ePt-643036* ePt-574367* ePt-573728*
3.1 ePt-568052
3.2 ePt-503896
4.7 Embra115
6.9 ePt-644409
7.8 ePt-571801*
7.9 ePt-571717
12.6 ePt-562757
14.5 Embra227
14.8 ePt-643823
23.9 ePt-638730
28.8 ePt-572244 ePt-641603
34.4 ePt-573853 ePt-575607 ePt-504479
35.0 ePt-640311
37.1 ePt-573996
37.2 ePt-641763
37.5 ePt-565337 ePt-569064
37.6 ePt-504002 ePt-570797 ePt-641758
37.7 ePt-599559 ePt-565694 ePt-571812
39.0 ePt-640491
39.3 ePt-570817 ePt-570900 ePt-573479
ePt-570469 ePt-600242
39.8 ePt-637193
40.0 ePt-573454 ePt-642290 ePt-567605
ePt-643636
40.1 ePt-639640
41.4 ePt-571733
44.6 ePt-566919
44.7 ePt-568396
44.8 ePt-574093
49.1 ePt-564735 ePt-642156
49.6 ePt-574190** ePt-569113
ePt-567264 ePt-564917 ePt-644403
49.8 ePt-644038 ePt-641733
49.9 ePt-640339
50.1 ePt-636969 ePt-638546 ePt-504090
ePt-572974 ePt-637164 ePt-600660
ePt-565850 ePt-643885 ePt-572614
ePt-643597 ePt-639702 ePt-572474
50.2 ePt-569172** ePt-644103** ePt-599887***
50.3 ePt-567194 ePt-567998 ePt-567849
50.8 ePt-504787
51.0 ePt-637279
53.6 ePt-639868* ePt-572802
53.7 ePt-599965* ePt-643259* ePt-600010*
59.8 ePt-573637
60.0 ePt-599305*
60.6 ePt-504168**
60.9 ePt-565565
61.1 ePt-639927
61.2 ePt-644491
61.6 ePt-636852
62.7 ePt-572180
62.9 ePt-639243
64.4 Embra1656
68.4 ePt-640855
70.8 ePt-564369
71.0 ePt-567027
75.1 ePt-568076**
75.7 ePt-641116** ePt-570139** ePt-565405**
83.4 ePt-600338*
83.6 ePt-644129*
86.8 ePt-571831*
89.4 ePt-566060*
89.5 ePt-600566*
89.7 ePt-575308
94.1 ePt-563549
99.8 ePt-639353 ePt-565169
99.9 ePt-639453 ePt-566453
101.2 ePt-637073 ePt-636581
101.4 ePt-574361
102.0 ePt-572573
102.7 ePt-640146

Figure S1.1A.

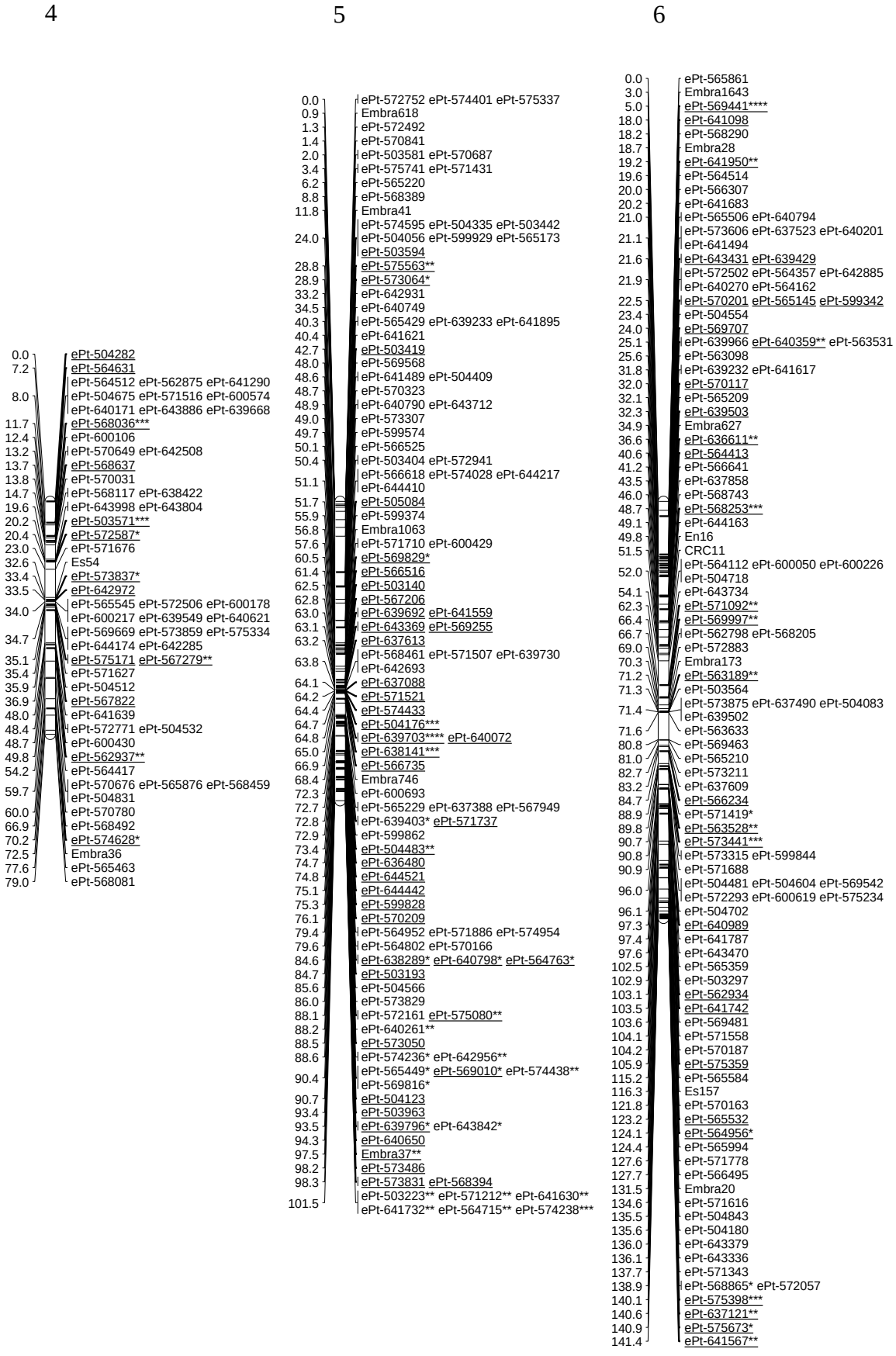
4
5
6
0.0 ePt-504282
7.2 ePt-564631
ePt-564512 ePt-562875 ePt-641290
8.0 ePt-504675 ePt-571516 ePt-600574
ePt-640171 ePt-643886 ePt-639668
11.7 ePt-568036***
12.4 ePt-600106
13.2 ePt-570649 ePt-642508
13.7 ePt-568637
13.8 ePt-570031
14.7 ePt-568117 ePt-638422
19.6 ePt-643998 ePt-643804
20.2 ePt-503571***
20.4 ePt-572587*
23.0 ePt-571676
32.6 Es54
33.4 ePt-573837*
33.5 ePt-642972
34.0 ePt-565545 ePt-572506 ePt-600178
ePt-600217 ePt-639549 ePt-640621
34.7 ePt-569669 ePt-573859 ePt-575334
ePt-644174 ePt-642285
35.1 ePt-575171 ePt-567279**
35.4 ePt-571627
35.9 ePt-504512
36.9 ePt-567822
48.0 ePt-641639
48.4 ePt-572771 ePt-504532
48.7 ePt-600430
49.8 ePt-562937**
54.2 ePt-564417
59.7 ePt-570676 ePt-565876 ePt-568459
60.0 ePt-504831
66.9 ePt-570780
70.2 ePt-568492
72.5 ePt-574628*
77.6 Embra36
79.0 ePt-565463
ePt-568081
0.0 ePt-572752 ePt-574401 ePt-575337
0.9 Embra618
1.3 ePt-572492
1.4 ePt-570841
2.0 ePt-503581 ePt-570687
3.4 ePt-575741 ePt-571431
6.2 ePt-565220
8.8 ePt-568389
11.8 Embra41
ePt-574595 ePt-504335 ePt-503442
24.0 ePt-504056 ePt-599929 ePt-565173
ePt-503594
28.8 ePt-575563**
28.9 ePt-573064*
33.2 ePt-642931
34.5 ePt-640749
40.3 ePt-565429 ePt-639233 ePt-641895
40.4 ePt-641621
42.7 ePt-503419
48.0 ePt-569568
48.6 ePt-641489 ePt-504409
48.7 ePt-570323
48.9 ePt-640790 ePt-643712
49.0 ePt-573307
49.7 ePt-599574
50.1 ePt-566525
50.4 ePt-503404 ePt-572941
51.1 ePt-566618 ePt-574028 ePt-644217
ePt-644410
51.7 ePt-505084
55.9 ePt-599374
56.8 Embra1063
57.6 ePt-571710 ePt-600429
60.5 ePt-569829*
61.4 ePt-566516
62.5 ePt-503140
62.8 ePt-567206
63.0 ePt-639692 ePt-641559
63.1 ePt-643369 ePt-569255
63.2 ePt-637613
63.8 ePt-568461 ePt-571507 ePt-639730
ePt-642693
64.1 ePt-637088
64.2 ePt-571521
64.4 ePt-574433
64.7 ePt-504176***
64.8 ePt-639703**** ePt-640072
65.0 ePt-638141***
66.9 ePt-566735
68.4 Embra746
72.3 ePt-600693
72.7 ePt-565229 ePt-637388 ePt-567949
72.8 ePt-639403* ePt-571737
72.9 ePt-599862
73.4 ePt-504483**
74.7 ePt-636480
74.8 ePt-644521
75.1 ePt-644442
75.3 ePt-599828
76.1 ePt-570209
79.4 ePt-564952 ePt-571886 ePt-574954
79.6 ePt-564802 ePt-570166
84.6 ePt-638289* ePt-640798* ePt-564763*
84.7 ePt-503193
85.6 ePt-504566
86.0 ePt-573829
88.1 ePt-572161 ePt-575080**
88.2 ePt-640261**
88.5 ePt-573050
88.6 ePt-574236* ePt-642956**
90.4 ePt-565449* ePt-569010* ePt-574438**
ePt-569816*
90.7 ePt-504123
93.4 ePt-503963
93.5 ePt-639796* ePt-643842*
94.3 ePt-640650
97.5 Embra37**
98.2 ePt-573486
98.3 ePt-573831 ePt-568394
101.5 ePt-503223** ePt-571212** ePt-641630**
ePt-641732** ePt-564715** ePt-574238***
0.0 ePt-565861
3.0 Embra1643
5.0 ePt-569441****
18.0 ePt-641098
18.2 ePt-568290
18.7 Embra28
19.2 ePt-641950**
19.6 ePt-564514
20.0 ePt-566307
20.2 ePt-641683
21.0 ePt-565506 ePt-640794
21.1 ePt-573606 ePt-637523 ePt-640201
ePt-641494
21.6 ePt-643431 ePt-639429
21.9 ePt-572502 ePt-564357 ePt-642885
ePt-640270 ePt-564162
22.5 ePt-570201 ePt-565145 ePt-599342
23.4 ePt-504554
24.0 ePt-569707
25.1 ePt-639966 ePt-640359** ePt-563531
25.6 ePt-563098
31.8 ePt-639232 ePt-641617
32.0 ePt-570117
32.1 ePt-565209
32.3 ePt-639503
34.9 Embra627
36.6 ePt-636611**
40.6 ePt-564413
41.2 ePt-566641
43.5 ePt-637858
46.0 ePt-568743
48.7 ePt-568253***
49.1 ePt-644163
49.8 En16
51.5 CRC11
52.0 ePt-564112 ePt-600050 ePt-600226
ePt-504718
54.1 ePt-643734
62.3 ePt-571092**
66.4 ePt-569997**
66.7 ePt-562798 ePt-568205
69.0 ePt-572883
70.3 Embra173
71.2 ePt-563189**
71.3 ePt-503564
71.4 ePt-573875 ePt-637490 ePt-504083
ePt-639502
71.6 ePt-563633
80.8 ePt-569463
81.0 ePt-565210
82.7 ePt-573211
83.2 ePt-637609
84.7 ePt-566234
88.9 ePt-571419*
89.8 ePt-563528**
90.7 ePt-573441***
90.8 ePt-573315 ePt-599844
90.9 ePt-571688
96.0 ePt-504481 ePt-504604 ePt-569542
ePt-572293 ePt-600619 ePt-575234
96.1 ePt-504702
97.3 ePt-640989
97.4 ePt-641787
97.6 ePt-643470
102.5 ePt-565359
102.9 ePt-503297
103.1 ePt-562934
103.5 ePt-641742
103.6 ePt-569481
104.1 ePt-571558
104.2 ePt-570187
105.9 ePt-575359
115.2 ePt-565584
116.3 Es157
121.8 ePt-570163
123.2 ePt-565532
124.1 ePt-564956*
124.4 ePt-565994
127.6 ePt-571778
127.7 ePt-566495
131.5 Embra20
134.6 ePt-571616
135.5 ePt-504843
135.6 ePt-504180
136.0 ePt-643379
136.1 ePt-643336
137.7 ePt-571343
138.9 ePt-568865* ePt-572057
140.1 ePt-575398***
140.6 ePt-637121**
140.9 ePt-575673*
141.4 ePt-641567**

Figure S1.1B.

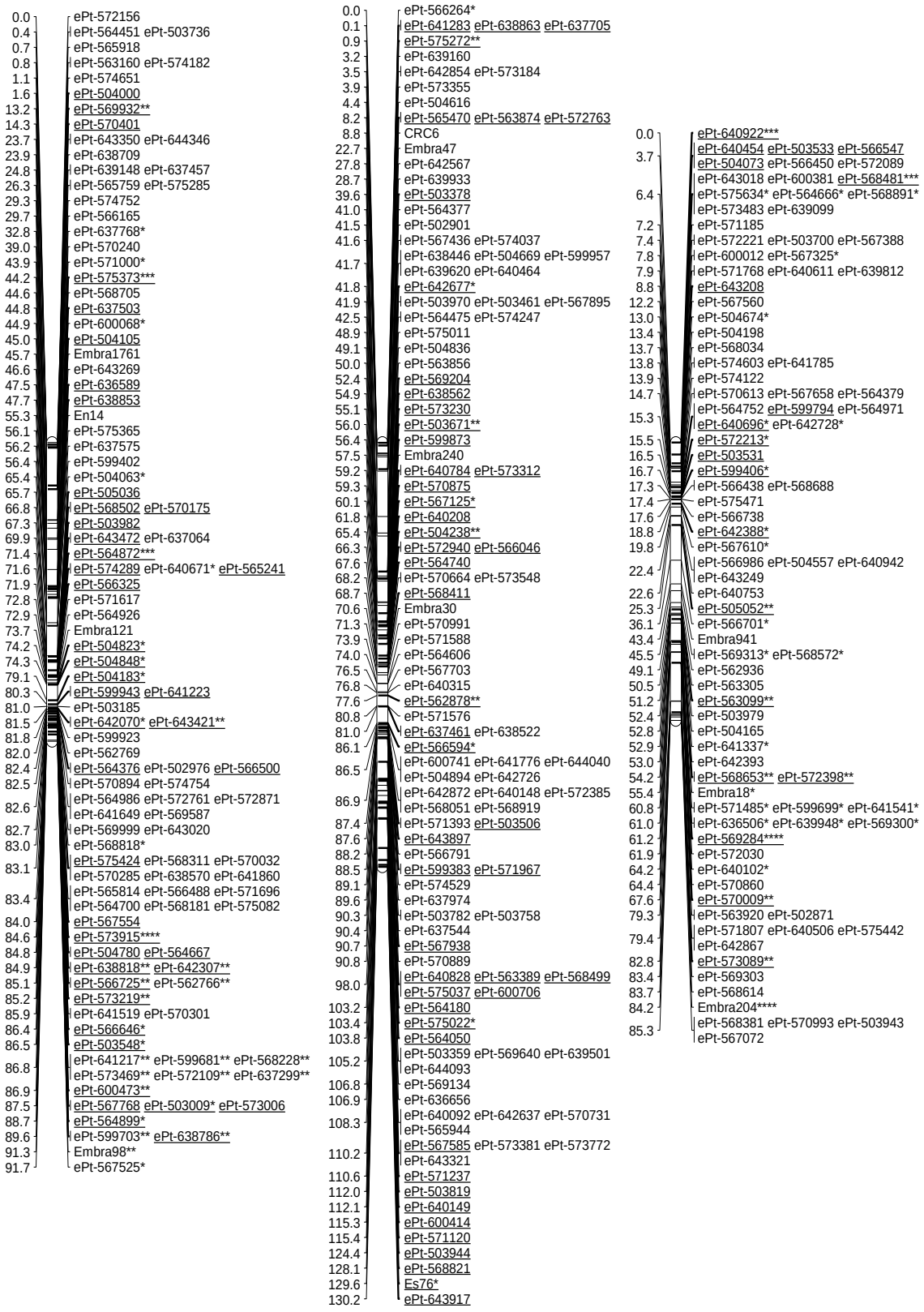
7
8
9
0.0 ePt-572156
0.4 ePt-564451 ePt-503736
0.7 ePt-565918
0.8 ePt-563160 ePt-574182
1.1 ePt-574651
1.6 ePt-504000
13.2 ePt-569932**
14.3 ePt-570401
23.7 ePt-643350 ePt-644346
23.9 ePt-638709
24.8 ePt-639148 ePt-637457
26.3 ePt-565759 ePt-575285
29.3 ePt-574752
29.7 ePt-566165
32.8 ePt-637768*
39.0 ePt-570240
43.9 ePt-571000*
44.2 ePt-575373***
44.6 ePt-568705
44.8 ePt-637503
44.9 ePt-600068*
45.0 ePt-504105
45.7 Embra1761
46.6 ePt-643269
47.5 ePt-636589
47.7 ePt-638853
55.3 En14
56.1 ePt-575365
56.2 ePt-637575
56.4 ePt-599402
65.4 ePt-504063*
65.7 ePt-505036
66.8 ePt-568502 ePt-570175
67.3 ePt-503982
69.9 ePt-643472 ePt-637064
71.4 ePt-564872***
71.6 ePt-574289 ePt-640671* ePt-565241
71.9 ePt-566325
72.8 ePt-571617
72.9 ePt-564926
73.7 Embra121
74.2 ePt-504823*
74.3 ePt-504848*
79.1 ePt-504183*
80.3 ePt-599943 ePt-641223
81.0 ePt-503185
81.5 ePt-642070* ePt-643421**
81.8 ePt-599923
82.0 ePt-562769
82.4 ePt-564376 ePt-502976 ePt-566500
82.5 ePt-570894 ePt-574754
82.6 ePt-564986 ePt-572761 ePt-572871
ePt-641649 ePt-569587
82.7 ePt-569999 ePt-643020
83.0 ePt-568818*
83.1 ePt-575424 ePt-568311 ePt-570032
ePt-570285 ePt-638570 ePt-641860
83.4 ePt-565814 ePt-566488 ePt-571696
ePt-564700 ePt-568181 ePt-575082
84.0 ePt-567554
84.6 ePt-573915****
84.8 ePt-504780 ePt-564667
84.9 ePt-638818** ePt-642307**
85.1 ePt-566725** ePt-562766**
85.2 ePt-573219**
85.9 ePt-641519 ePt-570301
86.4 ePt-566646*
86.5 ePt-503548*
86.8 ePt-641217** ePt-599681** ePt-568228**
ePt-573469** ePt-572109** ePt-637299**
86.9 ePt-600473**
87.5 ePt-567768 ePt-503009* ePt-573006
88.7 ePt-564899*
89.6 ePt-599703** ePt-638786**
91.3 Embra98**
91.7 ePt-567525*
0.0 ePt-566264*
0.1 ePt-641283 ePt-638863 ePt-637705
0.9 ePt-575272**
3.2 ePt-639160
3.5 ePt-642854 ePt-573184
3.9 ePt-573355
4.4 ePt-504616
8.2 ePt-565470 ePt-563874 ePt-572763
8.8 CRC6
22.7 Embra47
27.8 ePt-642567
28.7 ePt-639933
39.6 ePt-503378
41.0 ePt-564377
41.5 ePt-502901
41.6 ePt-567436 ePt-574037
41.7 ePt-638446 ePt-504669 ePt-599957
ePt-639620 ePt-640464
41.8 ePt-642677*
41.9 ePt-503970 ePt-503461 ePt-567895
42.5 ePt-564475 ePt-574247
48.9 ePt-575011
49.1 ePt-504836
50.0 ePt-563856
52.4 ePt-569204
54.9 ePt-638562
55.1 ePt-573230
56.0 ePt-503671**
56.4 ePt-599873
57.5 Embra240
59.2 ePt-640784 ePt-573312
59.3 ePt-570875
60.1 ePt-567125*
61.8 ePt-640208
65.4 ePt-504238**
66.3 ePt-572940 ePt-566046
67.6 ePt-564740
68.2 ePt-570664 ePt-573548
68.7 ePt-568411
70.6 Embra30
71.3 ePt-570991
73.9 ePt-571588
74.0 ePt-564606
76.5 ePt-567703
76.8 ePt-640315
77.6 ePt-562878**
80.8 ePt-571576
81.0 ePt-637461 ePt-638522
86.1 ePt-566594*
86.5 ePt-600741 ePt-641776 ePt-644040
ePt-504894 ePt-642726
ePt-642872 ePt-640148 ePt-572385
86.9 ePt-568051 ePt-568919
87.4 ePt-571393 ePt-503506
87.6 ePt-643897
88.2 ePt-566791
88.5 ePt-599383 ePt-571967
89.1 ePt-574529
89.6 ePt-637974
90.3 ePt-503782 ePt-503758
90.4 ePt-637544
90.7 ePt-567938
90.8 ePt-570889
98.0 ePt-640828 ePt-563389 ePt-568499
ePt-575037 ePt-600706
103.2 ePt-564180
103.4 ePt-575022*
103.8 ePt-564050
105.2 ePt-503359 ePt-569640 ePt-639501
ePt-644093
106.8 ePt-569134
106.9 ePt-636656
108.3 ePt-640092 ePt-642637 ePt-570731
ePt-565944
110.2 ePt-567585 ePt-573381 ePt-573772
ePt-643321
110.6 ePt-571237
112.0 ePt-503819
112.1 ePt-640149
115.3 ePt-600414
115.4 ePt-571120
124.4 ePt-503944
128.1 ePt-568821
129.6 Es76*
130.2 ePt-643917
0.0 ePt-640922***
3.7 ePt-640454 ePt-503533 ePt-566547
ePt-504073 ePt-566450 ePt-572089
6.4 ePt-643018 ePt-600381 ePt-568481***
ePt-575634* ePt-564666* ePt-568891*
ePt-573483 ePt-639099
7.2 ePt-571185
7.4 ePt-572221 ePt-503700 ePt-567388
7.8 ePt-600012 ePt-567325*
7.9 ePt-571768 ePt-640611 ePt-639812
8.8 ePt-643208
12.2 ePt-567560
13.0 ePt-504674*
13.4 ePt-504198
13.7 ePt-568034
13.8 ePt-574603 ePt-641785
13.9 ePt-574122
14.7 ePt-570613 ePt-567658 ePt-564379
ePt-564752 ePt-599794 ePt-564971
15.3 ePt-640696* ePt-642728*
15.5 ePt-572213*
16.5 ePt-503531
16.7 ePt-599406*
17.3 ePt-566438 ePt-568688
17.4 ePt-575471
17.6 ePt-566738
18.8 ePt-642388*
19.8 ePt-567610*
22.4 ePt-566986 ePt-504557 ePt-640942
ePt-643249
22.6 ePt-640753
25.3 ePt-505052**
36.1 ePt-566701*
43.4 Embra941
45.5 ePt-569313* ePt-568572*
49.1 ePt-562936
50.5 ePt-563305
51.2 ePt-563099**
52.4 ePt-503979
52.8 ePt-504165
52.9 ePt-641337*
53.0 ePt-642393
54.2 ePt-568653** ePt-572398**
55.4 Embra18*
60.8 ePt-571485* ePt-599699* ePt-641541*
61.0 ePt-636506* ePt-639948* ePt-569300*
61.2 ePt-569284****
61.9 ePt-572030
64.2 ePt-640102*
64.4 ePt-570860
67.6 ePt-570009**
79.3 ePt-563920 ePt-502871
79.4 ePt-571807 ePt-640506 ePt-575442
ePt-642867
82.8 ePt-573089**
83.4 ePt-569303
83.7 ePt-568614
84.2 Embra204****
85.3 ePt-568381 ePt-570993 ePt-503943
ePt-567072

Figure S1.1C.

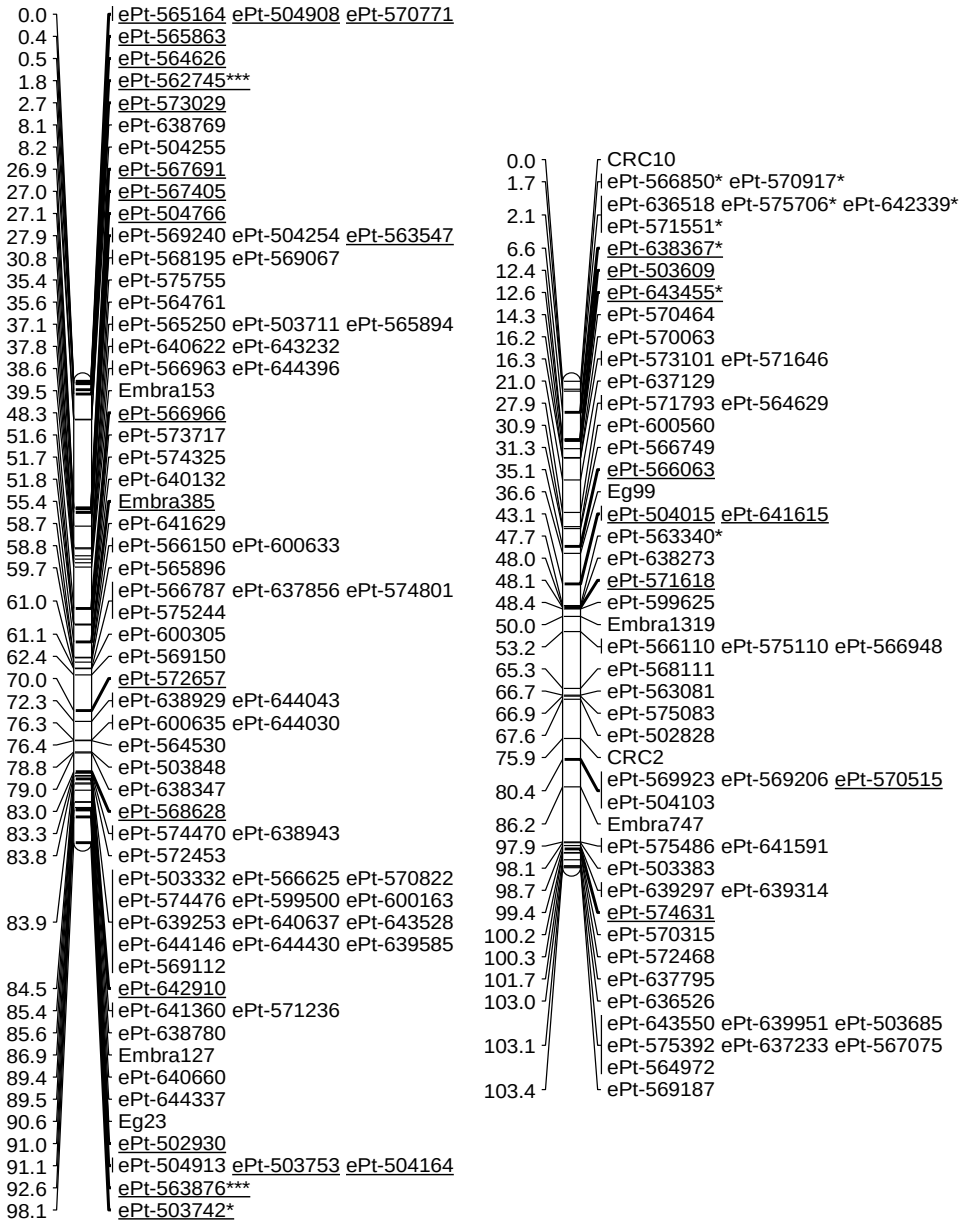
10
11
0.0 ePt-565164 ePt-504908 ePt-570771
0.4 ePt-565863
0.5 ePt-564626
1.8 ePt-562745***
2.7 ePt-573029
8.1 ePt-638769
8.2 ePt-504255
26.9 ePt-567691
27.0 ePt-567405
27.1 ePt-504766
27.9 ePt-569240 ePt-504254 ePt-563547
30.8 ePt-568195 ePt-569067
35.4 ePt-575755
35.6 ePt-564761
37.1 ePt-565250 ePt-503711 ePt-565894
37.8 ePt-640622 ePt-643232
38.6 ePt-566963 ePt-644396
39.5 Embra153
48.3 ePt-566966
51.6 ePt-573717
51.7 ePt-574325
51.8 ePt-640132
55.4 Embra385
58.7 ePt-641629
58.8 ePt-566150 ePt-600633
59.7 ePt-565896
61.0 ePt-566787 ePt-637856 ePt-574801
ePt-575244
61.1 ePt-600305
62.4 ePt-569150
70.0 ePt-572657
72.3 ePt-638929 ePt-644043
76.3 ePt-600635 ePt-644030
76.4 ePt-564530
78.8 ePt-503848
79.0 ePt-638347
83.0 ePt-568628
83.3 ePt-574470 ePt-638943
83.8 ePt-572453
83.9 ePt-503332 ePt-566625 ePt-570822
ePt-574476 ePt-599500 ePt-600163
ePt-639253 ePt-640637 ePt-643528
ePt-644146 ePt-644430 ePt-639585
ePt-569112
84.5 ePt-642910
85.4 ePt-641360 ePt-571236
85.6 ePt-638780
86.9 Embra127
89.4 ePt-640660
89.5 ePt-644337
90.6 Eg23
91.0 ePt-502930
91.1 ePt-504913 ePt-503753 ePt-504164
92.6 ePt-563876***
98.1 ePt-503742*
0.0 CRC10
1.7 ePt-566850* ePt-570917*
2.1 ePt-636518 ePt-575706* ePt-642339*
ePt-571551*
6.6 ePt-638367*
12.4 ePt-503609
12.6 ePt-643455*
14.3 ePt-570464
16.2 ePt-570063
16.3 ePt-573101 ePt-571646
21.0 ePt-637129
27.9 ePt-571793 ePt-564629
30.9 ePt-600560
31.3 ePt-566749
35.1 ePt-566063
36.6 Eg99
43.1 ePt-504015 ePt-641615
47.7 ePt-563340*
48.0 ePt-638273
48.1 ePt-571618
48.4 ePt-599625
50.0 Embra1319
53.2 ePt-566110 ePt-575110 ePt-566948
65.3 ePt-568111
66.7 ePt-563081
66.9 ePt-575083
67.6 ePt-502828
75.9 CRC2
80.4 ePt-569923 ePt-569206 ePt-570515
ePt-504103
86.2 Embra747
97.9 ePt-575486 ePt-641591
98.1 ePt-503383
98.7 ePt-639297 ePt-639314
99.4 ePt-574631
100.2 ePt-570315
100.3 ePt-572468
101.7 ePt-637795
103.0 ePt-636526
103.1 ePt-643550 ePt-639951 ePt-503685
ePt-575392 ePt-637233 ePt-567075
ePt-564972
103.4 ePt-569187

Figure S1.1D.

This page and following pages:

Figure S2.1A-L Marker colinearity between all six component maps and the *Eucalyptus* multi-species composite map. For LG4, comparisons between both GU-SA and GLOB-LH seed-map composite maps are shown. For each LG, three LG 'triplets' show marker colinearity between two component maps (outside) and the composite map (centre). Horizontal lines on LG bars indicate marker positions and lines between LGs indicate the position of common markers. The scale bar shown is in Kosambi's centiMorgans. Component map names (abbreviations; refer to Table 2.1) are given above each LG. LGs excluded from composite map construction are indicated in parentheses following the component map name. An asterisk indicates whether marker-order information from the component map was incorporated during composite map construction (see material and methods Chapter 2). For the GU-Emb component map, superscript letters indicates whether the framework (f) or comprehensive (c) linkage group from this pedigree was used in composite map construction.

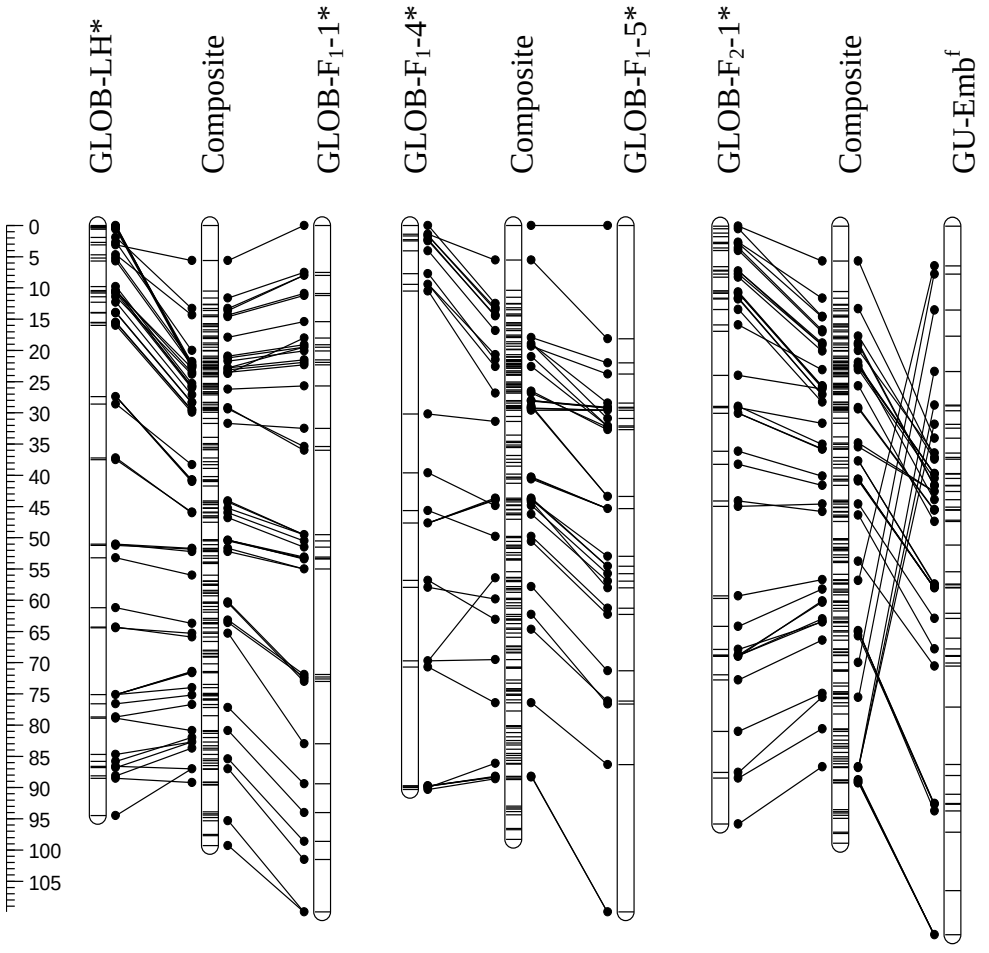

Figure S2.1A Linkage group 1.

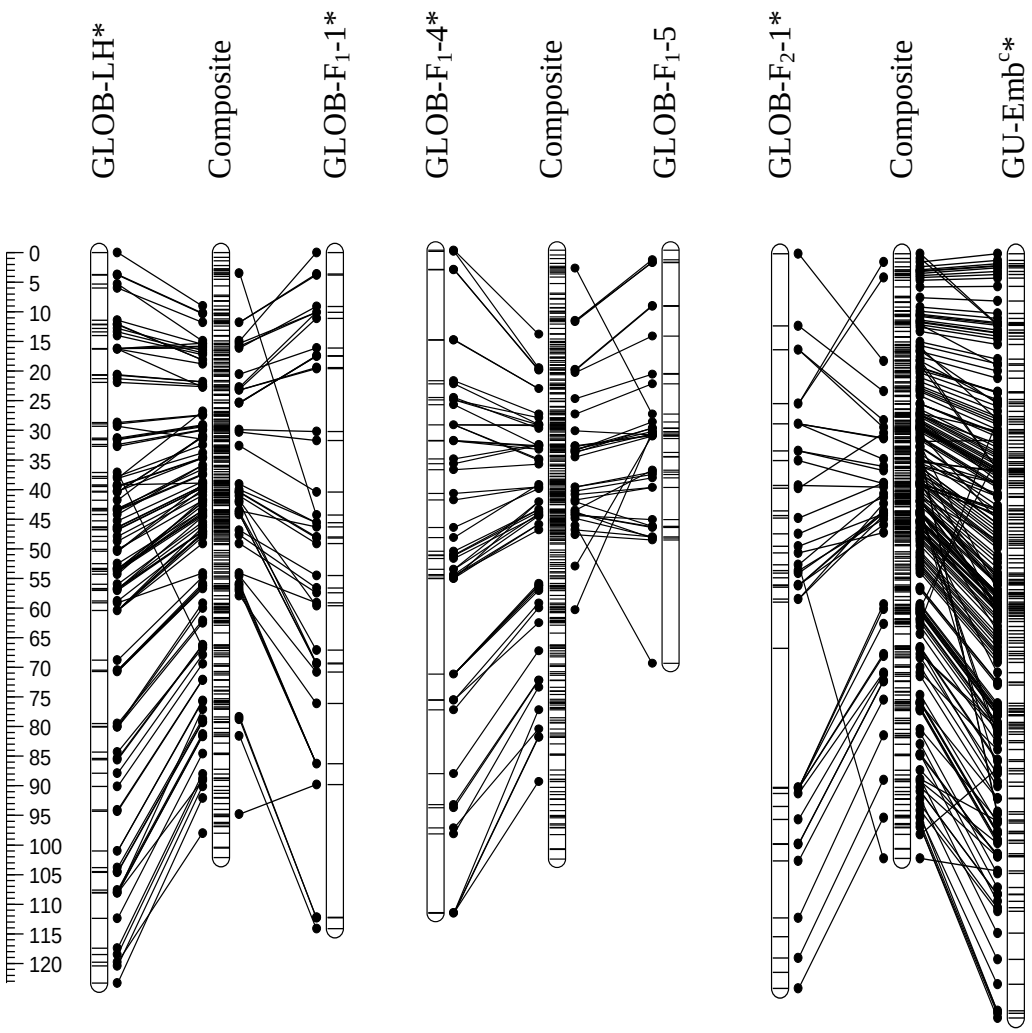

Figure S2.1B Linkage group 2.

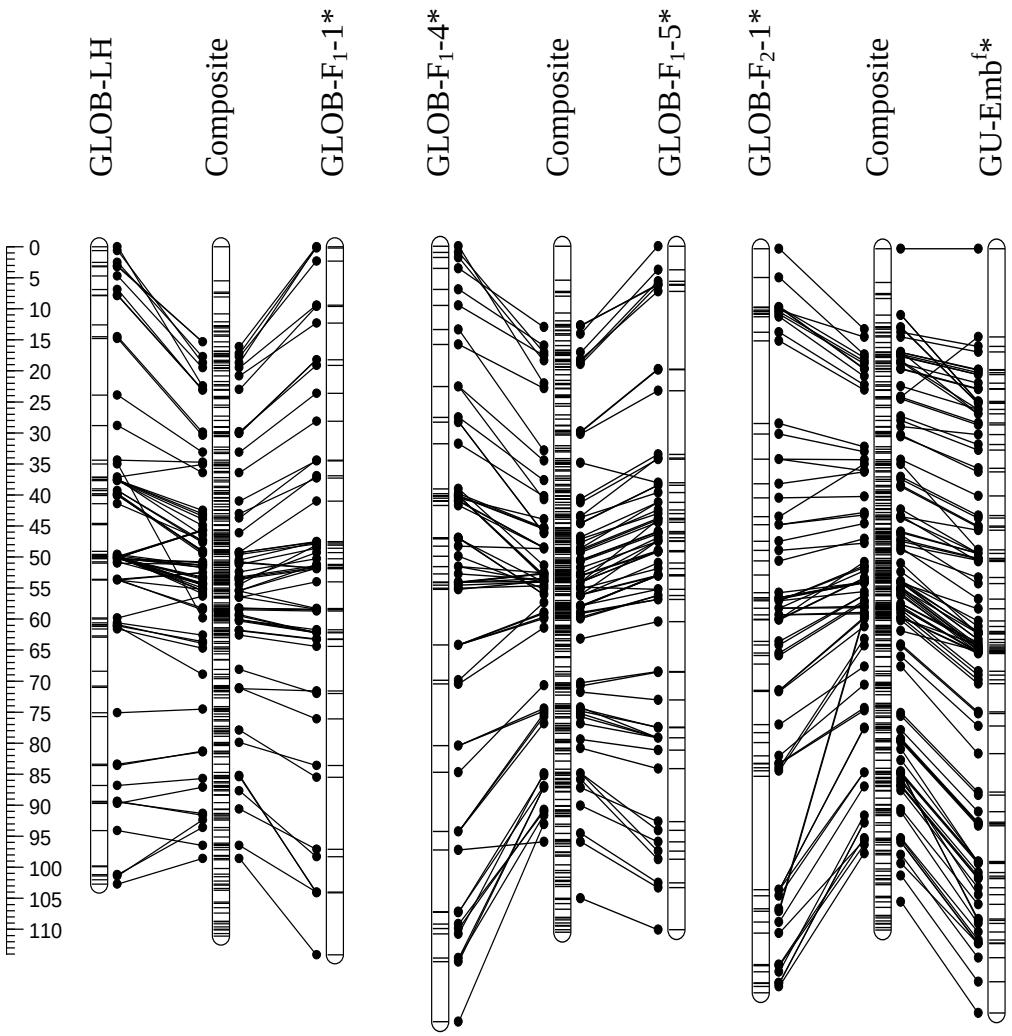

Figure S2.1C Linkage group 3.

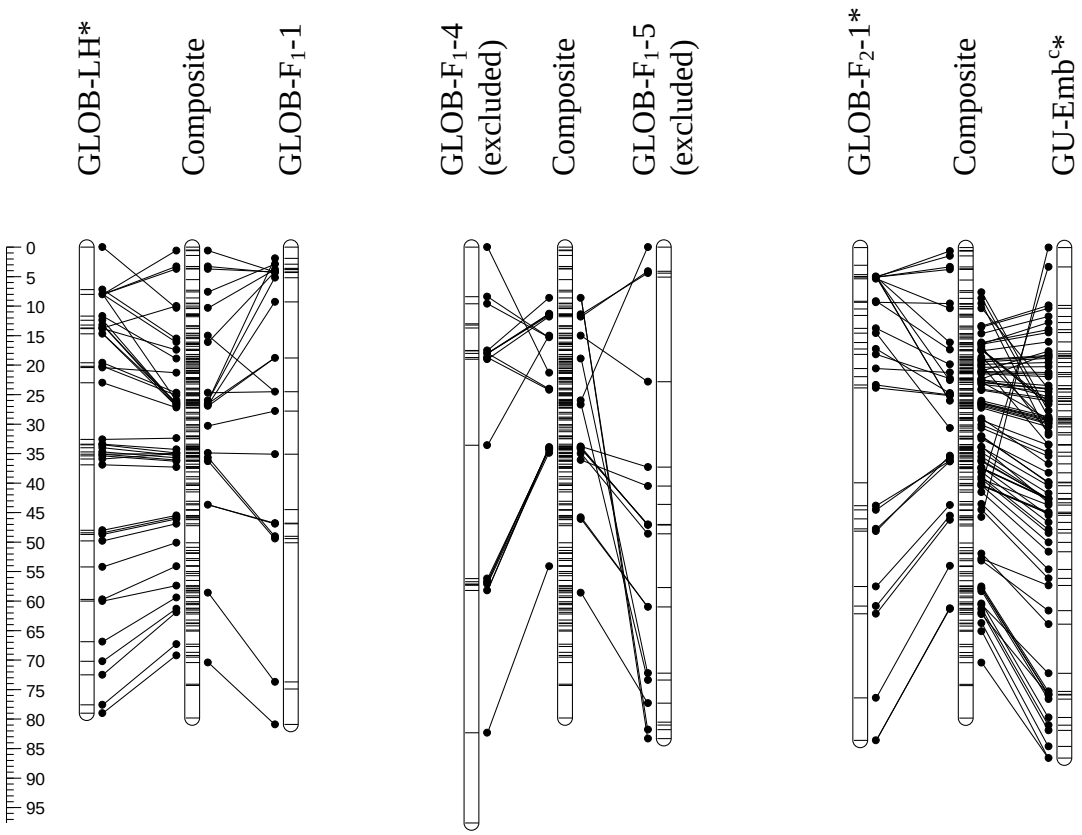

Figure S2.1D Linkage group 4 (GU-SA seed-map).

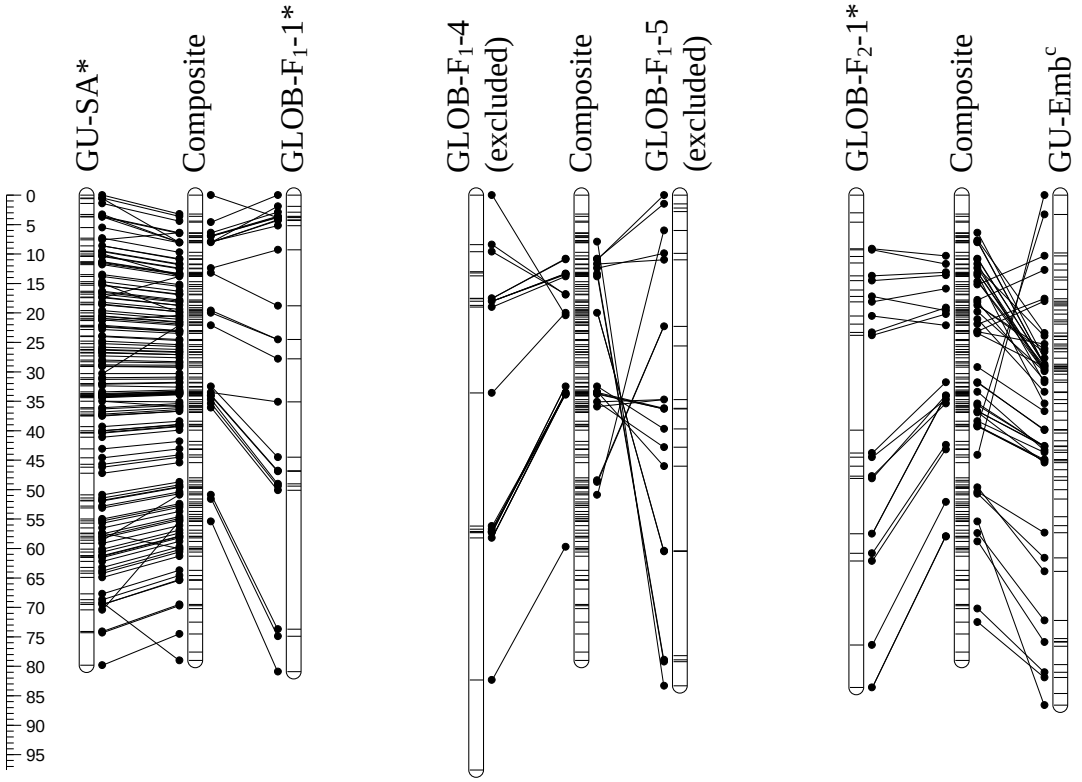

Figure S2.1E Linkage group 4 (GLOB-LH seed-map). Linkage group 4 built using the GLOB-LH seed-map was included in the *Eucalyptus* multi-species composite map.

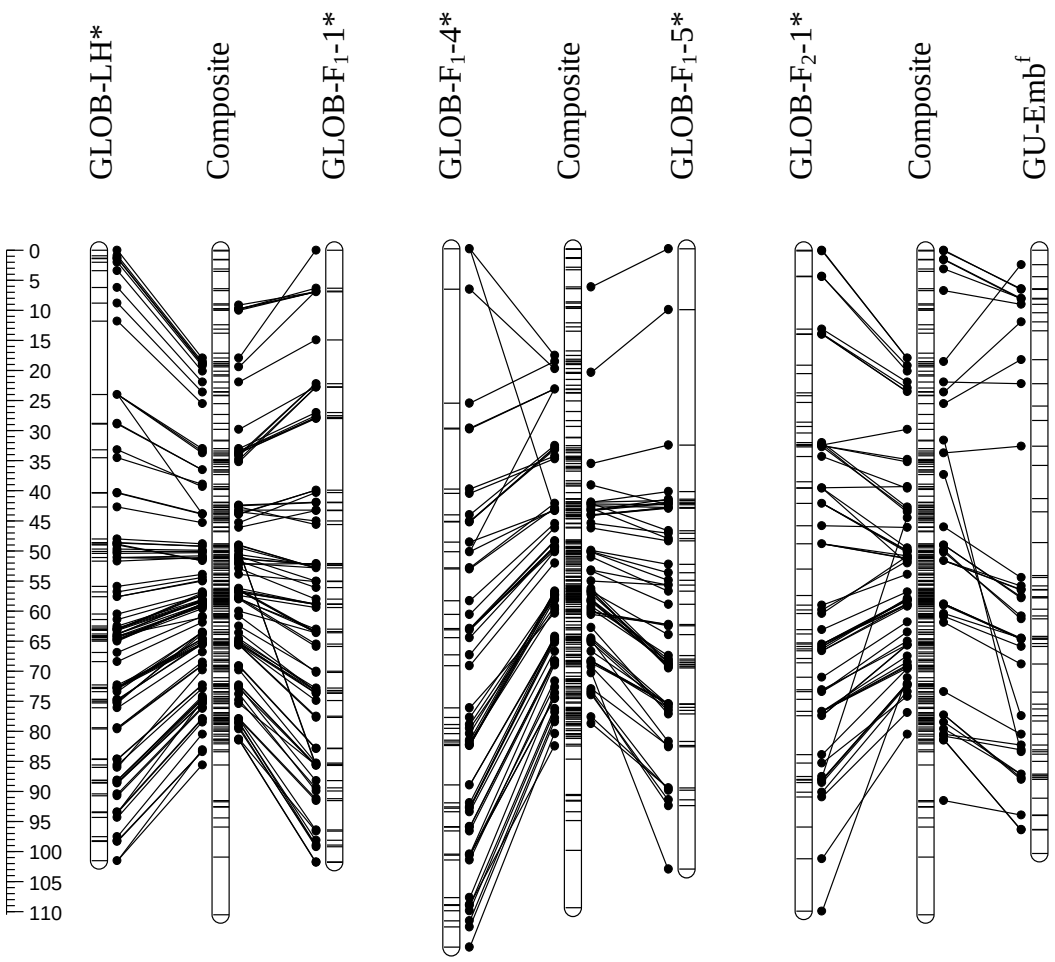

Figure S2.1F Linkage group 5.

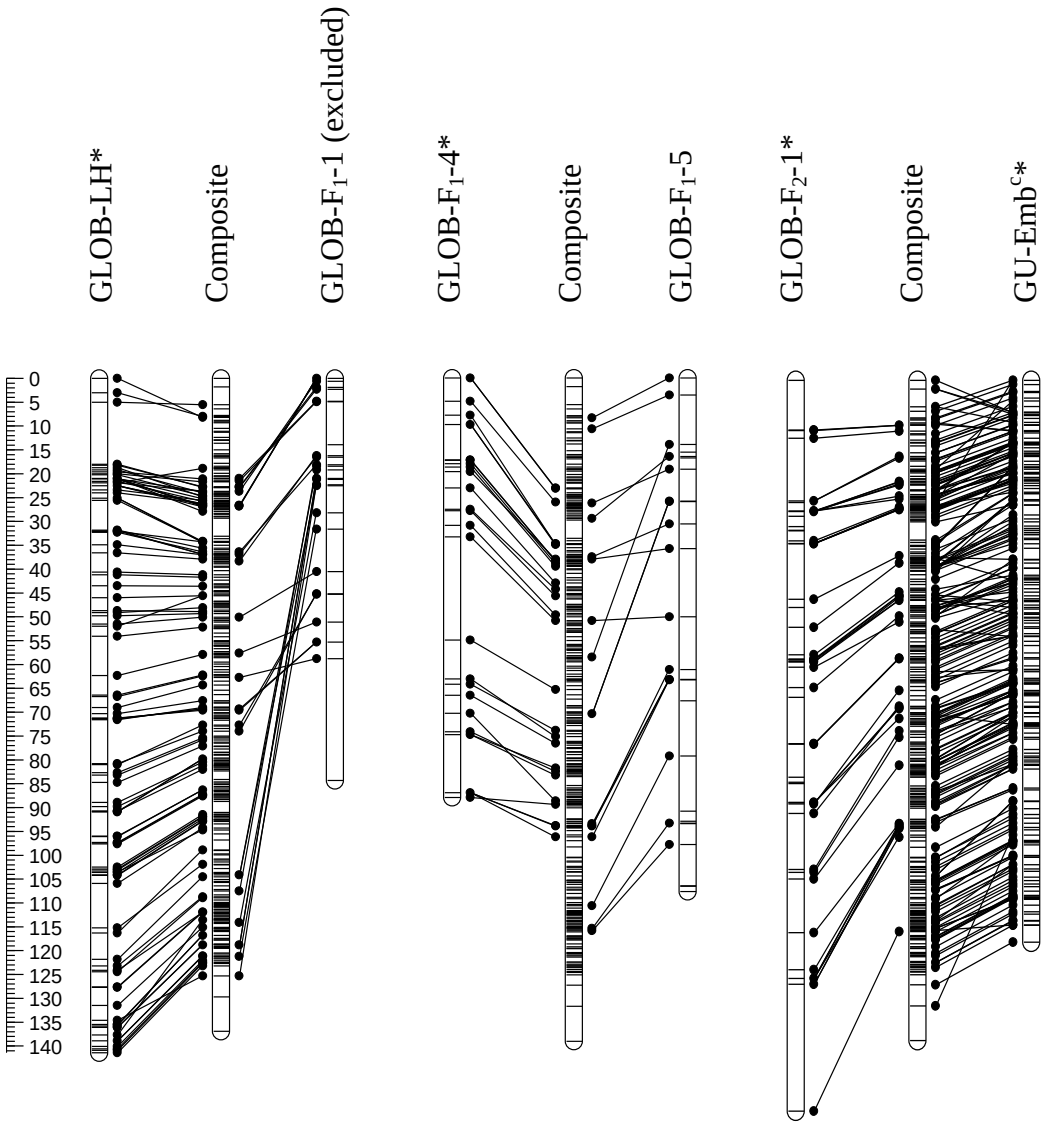

Figure S2.1G Linkage group 6.

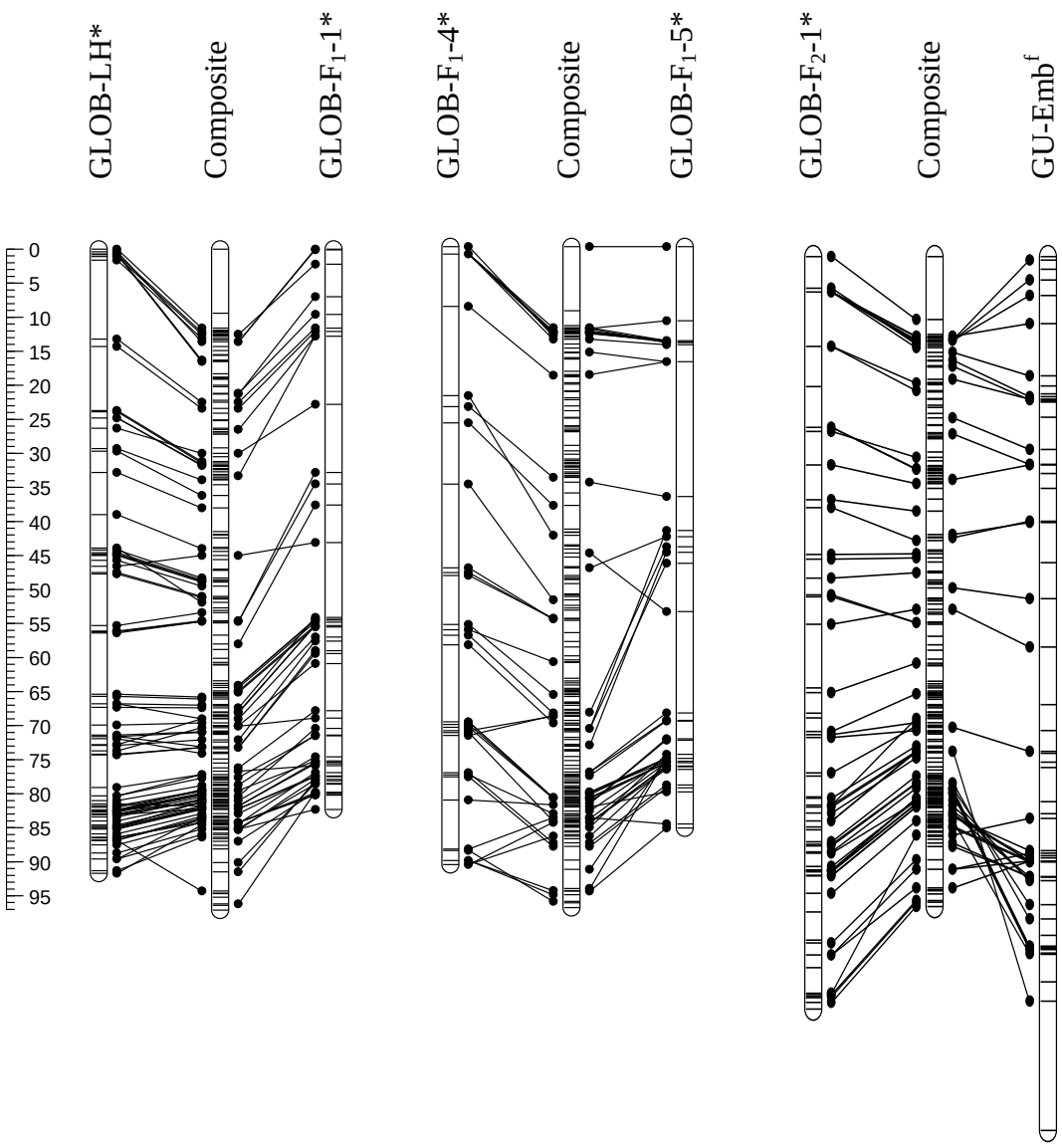

Figure S2.1H Linkage group 7.

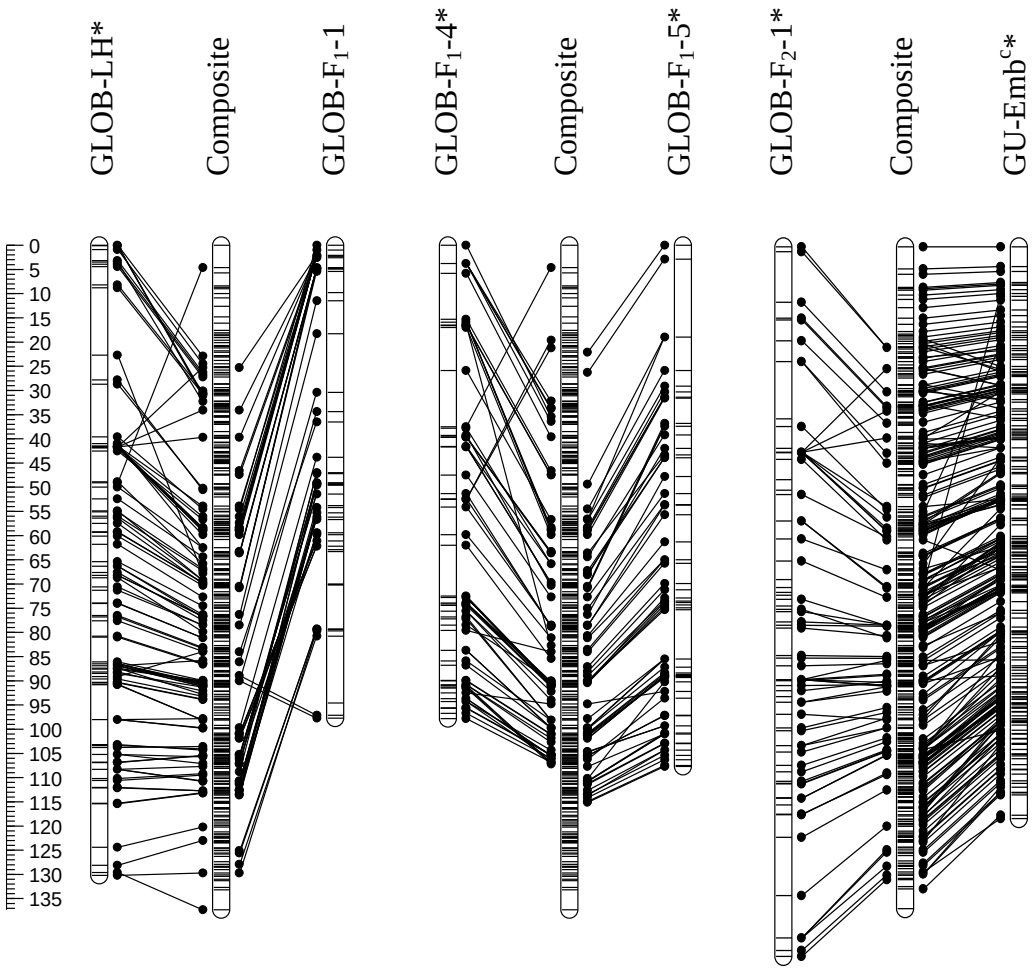

Figure S2.1I Linkage group 8.

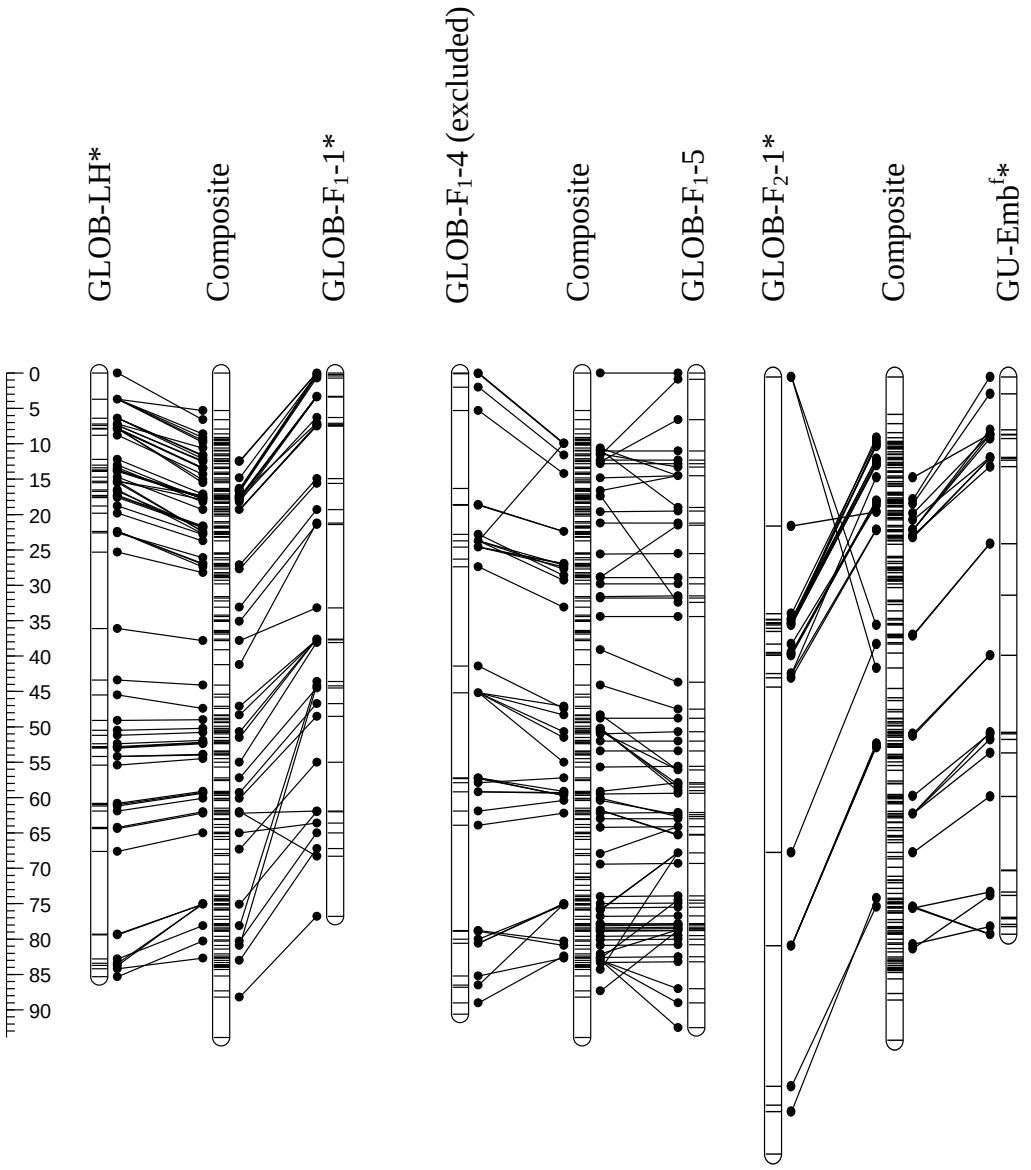

Figure S2.1J Linkage group 9.

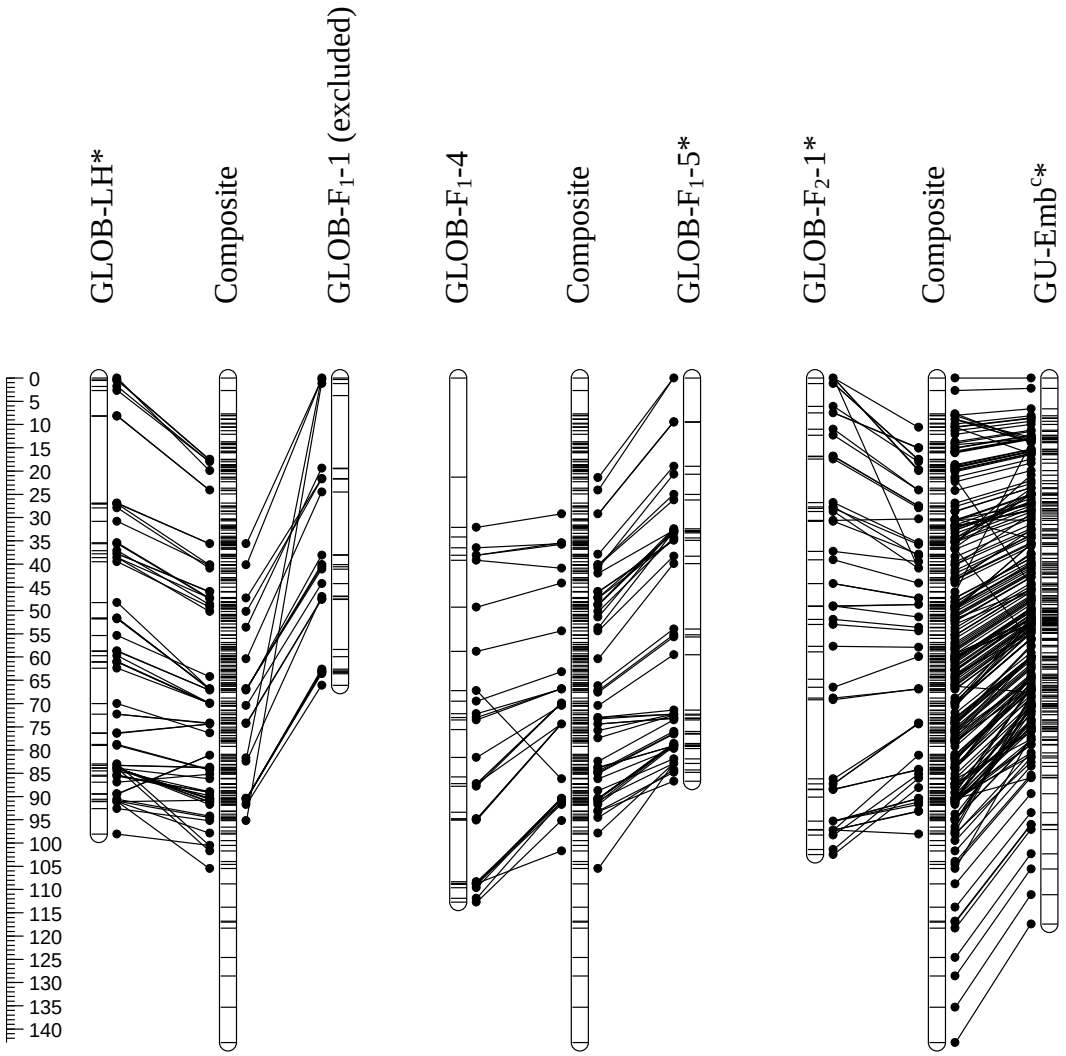

Figure S2.1K Linkage group 10.

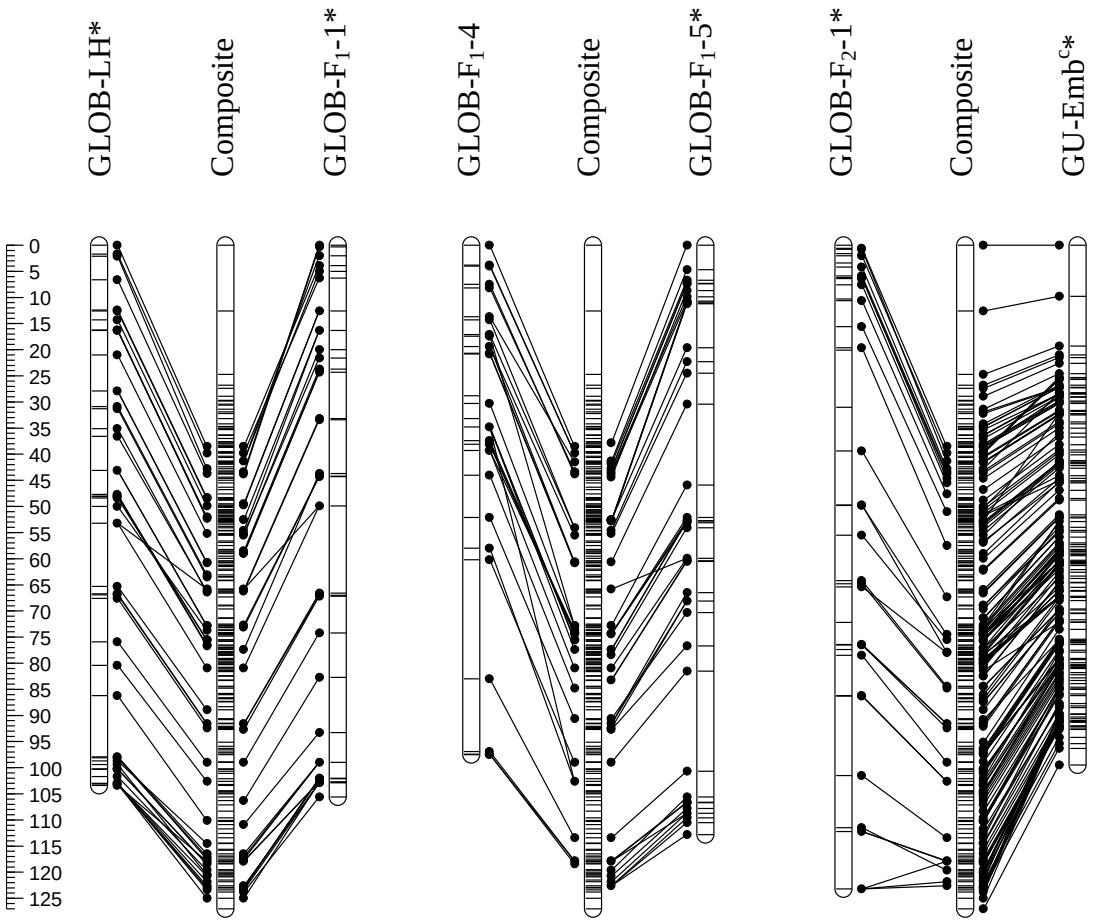

Figure S2.1L Linkage group 11.

This page and following pages: Figure S2.2A-C Comparison of linkage group marker-orders between *Eucalyptus* multi-species composite maps built using different seed-maps; GU-SA seed-map (left) and GLOB-LH seed-map (right). Horizontal lines on linkage group bars indicate marker positions and lines between linkage groups indicate the position of common markers. Scale bar is in Kosambi's centiMorgans.

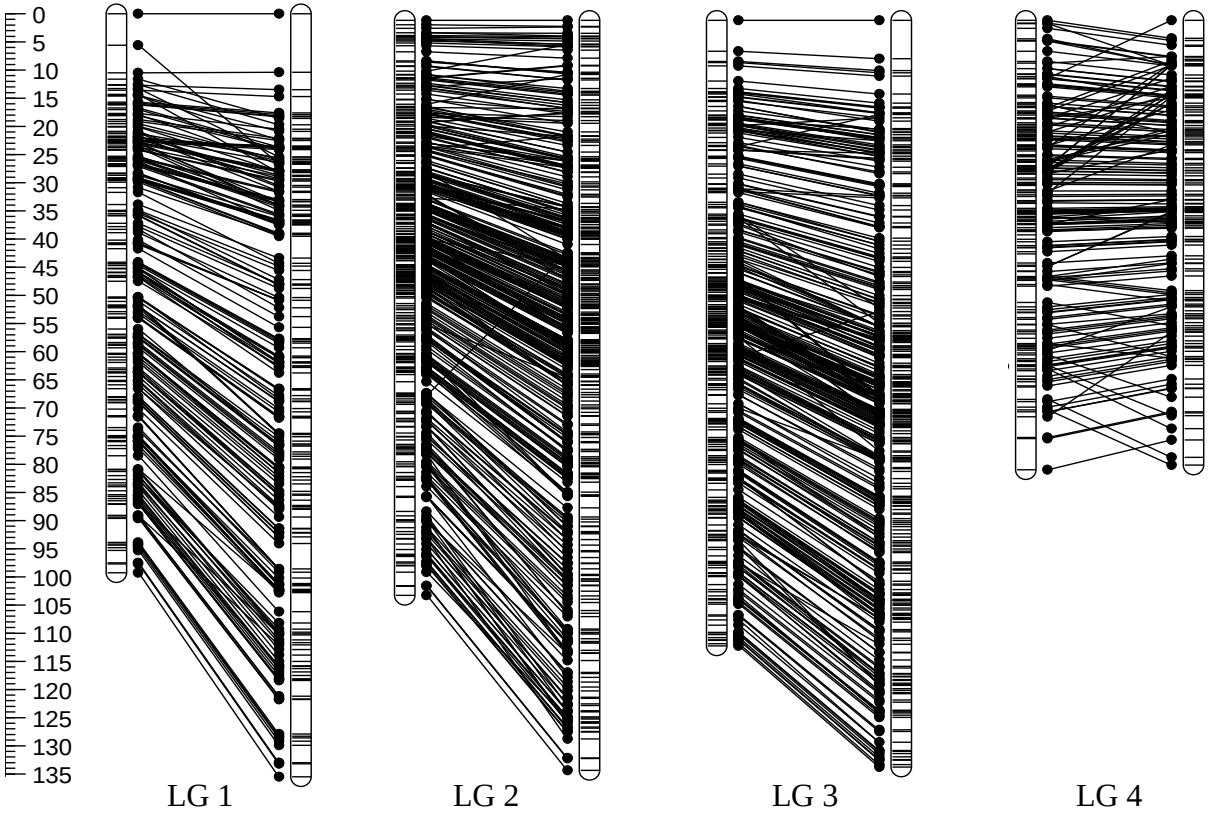

Figure S2.2A Comparison of GU-SA seed-map (left) and GLOB-LH seed-map (right) composite map linkage groups 1 to 4.

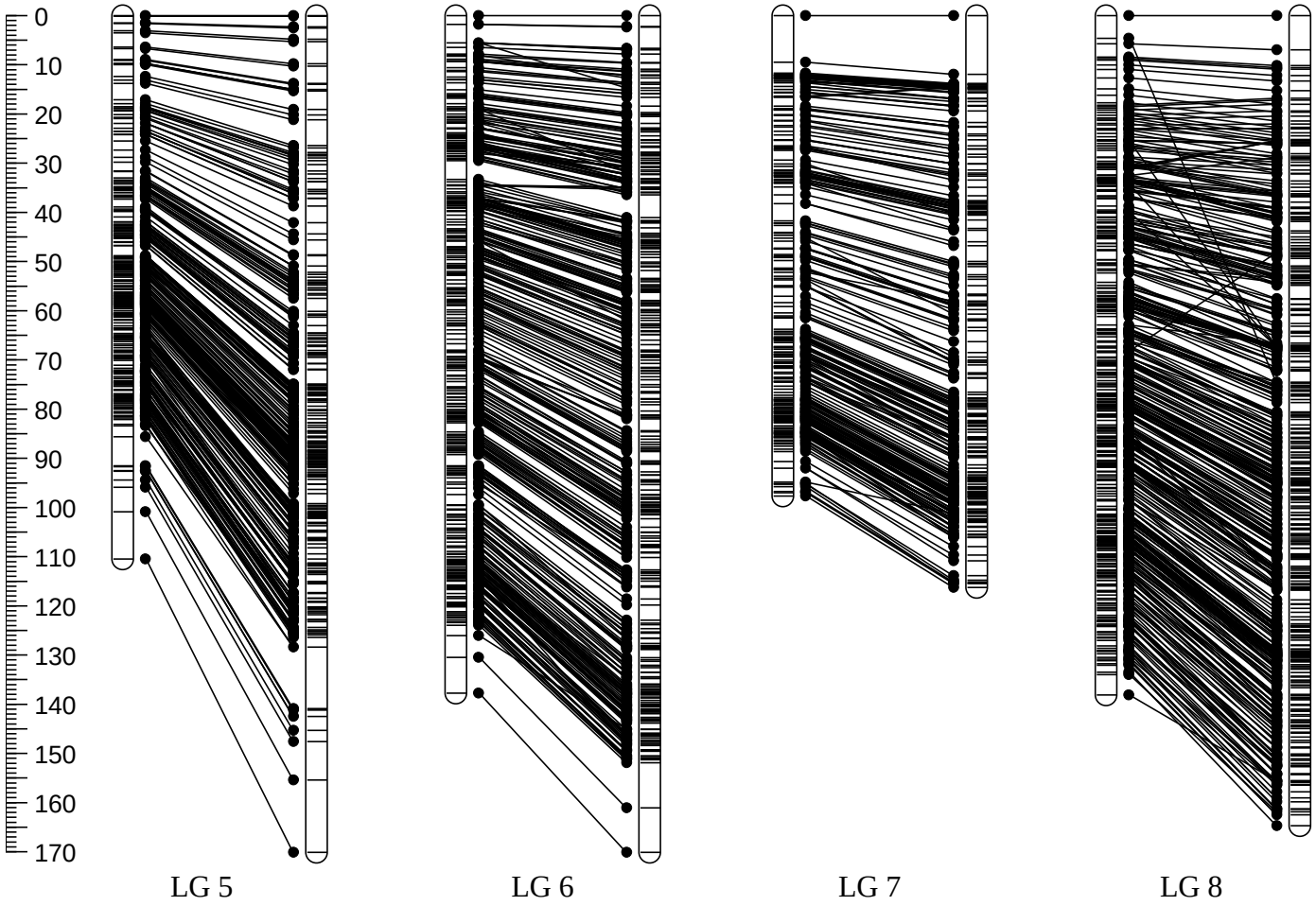

Figure S2.2B Comparison of GU-SA seed-map (left) and GLOB-LH seed-map (right) composite map linkage groups 5 to 8.

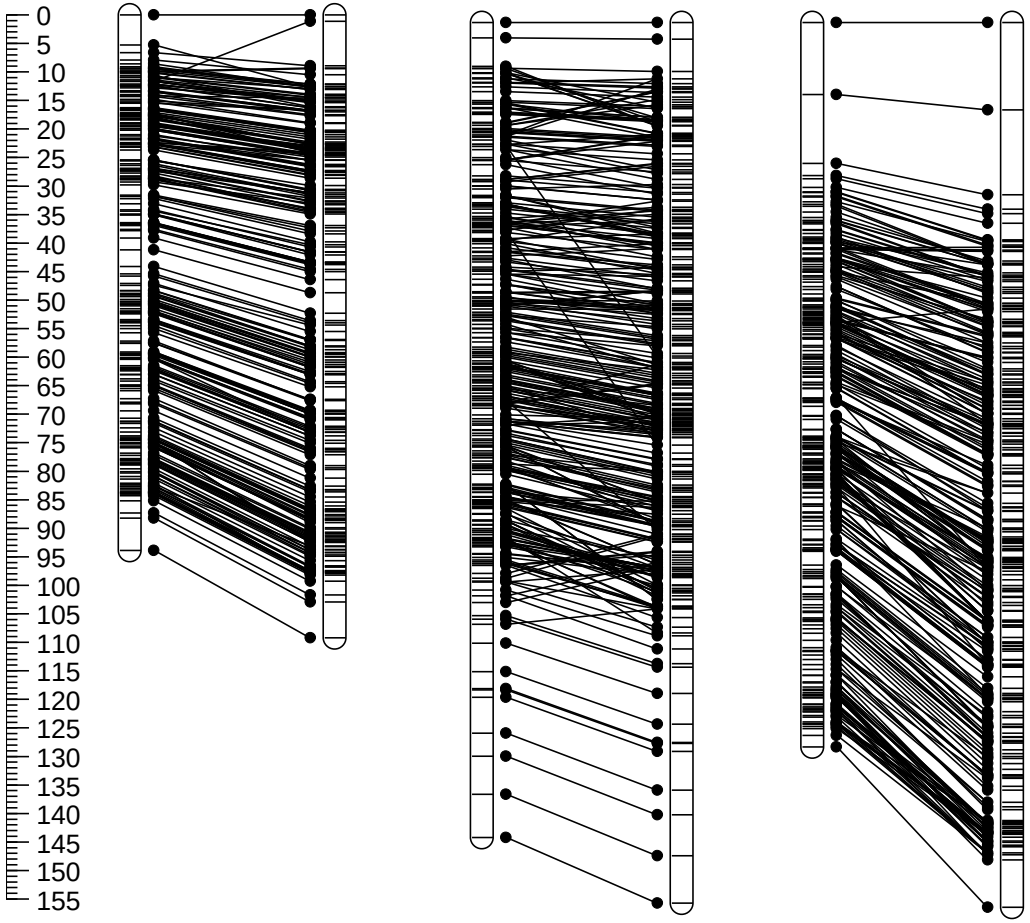

Figure S2.2C Comparison of GU-SA seed-map (left) and GLOB-LH seed-map (right) composite map linkage groups 9 to 11.

References

Achere V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology* **14**, 3191-3201.

Agrama HA, Salah SF (2002) Construction of a genome map for *Eucalyptus camaldulensis* DEHN. *Silvae Genetica* **51**, 201-206.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* **12**, 1805-1814.

Alsop B, Farre A, Wenzl P, Wang J, Zhou M, Romagosa I, Kilian A, Steffenson B (2011) Development of wild barley-derived DArT markers and their integration into a barley consensus map. *Molecular Breeding* **27**, 77-92.

Anderson JT, Lee C-R, Mitchell-Olds T (2011) Life-history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*. *Evolution* **65**, 771-87.

Asíns MJ, Mestre P, García JE, Dicenta F, Carbonell EA (1994) Genotype x environment interaction in QTL analysis of an intervarietal almond cross by means of genetic markers. *Theoretical and Applied Genetics* **89**, 358-364.

Bachir O, Abdellah B (2006) Chromosome numbers of the 59 species of *Eucalyptus* L'Herit (*Myrtaceae*). *Caryologia* **59**, 207-212.

Bäurle I, Dean C (2006) The timing of developmental transitions in plants. *Cell* **125**, 655-664.

Bedo J, Wenzl P, Kowalczyk A, Kilian A (2008) Precision-mapping and statistical validation of quantitative trait loci by machine learning. *BMC Genetics* **9:35**.

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *The Plant Cell* **12**, 1021-1029.

Bennetzen JL (2007) Patterns in grass genome evolution. *Current Opinion in Plant Biology* **10**, 176-181.

Bennetzen JL, Ma J, Devos K (2005) Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**, 127-132.

Bennington CC, McGraw JB (1995) Natural selection and ecotypic differentiation in *Impatiens Pallida*. *Ecological Monographs* **65**, 303-323.

Bergonzi S, Albani MC (2011) Reproductive competence from an annual and a perennial perspective. *Journal of Experimental Botany*, doi:10.1093/jxb/err192.

Bodenes C, Joandet S, Laigret F, Kremer A (1997) Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Heredity* **78**, 433-444.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* **23**, 773-783.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet* **6**, e1000940.

Bratteler M, Baltisbeger M, Widmner A (2006) QTL analysis of intraspecific differences between two *Silene vulgaris* ecotypes. *Annals of Botany* **98**, 411-419.

Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* **267**, 338-347.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* **97**, 816-827.

Brondani RPV, Williams ER, Brondani C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* **6:20**.

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (*Myrtaceae*). *Australian Systematic Botany* **13**, 79-148.

Brooker MIH, Kleinig DA (1999) 'Field guide to eucalypts. Vol. 1, South-eastern Australia.' (Bloomings Books: Melbourne)

Brooker MIH, Kleinig DA (2004) 'Field guide to eucalypts. Vol. 3, Northern Australia.' (Bloomings Books: Melbourne)

Brown GR, Bassoni DL, *et al.* (2003) Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics* **164**, 1537-1546.

Bundock P, Potts B, Vaillancourt R (2008) Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. *Tree Genetics & Genomes* **4**, 85-95.

Butcher P, McDonald M, Bell J (2009) Congruence between environmental parameters, morphology and genetic structure in Australia's most widely distributed eucalypt, *Eucalyptus camaldulensis*. *Tree Genetics & Genomes* **5**, 189-210.

Butlin R (2010) Population genomics and speciation. *Genetica* **138**, 409-418.

Byrne M (2008) Phylogeny, diversity and evolution of eucalypts. In 'Plant Genome - Biodiversity and Evolution'. (Eds AK Sharma and A Sharma) pp. 303-346. (Science Publishers: Enfield NH)

Byrne M, Marquezgarcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Australian Journal of Botany* **44**, 331-341.

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. *Theoretical and Applied Genetics* **91**, 869-875.

Byrne M, Murrell JC, Owen JV, Williams ER, Moran GF (1997) Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theoretical and Applied Genetics* **95**, 975-979.

Byrne M, Parrish TL, Moran GF (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. *Heredity* **81**, 225-233.

Carlsbecker A, Tandre K, Johanson U, Englund M, Engström P (2004) The MADS-box gene *DAL1* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *The Plant Journal* **40**, 546-557.

Carver E, Stubbs L (1997) Zooming in on the human-mouse comparative map: Genome conservation re-examined on a high-resolution scale. *Genome Research* **7**, 1123-1137.

Celton J-M, Chagne D, Tustin S, Terakami S, Nishitani C, Yamamoto T, Gardiner S (2009) Update on comparative genome mapping between *Malus* and *Pyrus*. *BMC Research Notes* **2**, 182-188.

Chalmers PM (1992) The adaptive significance of juvenile versus adult leaves in *Eucalyptus globulus* ssp. *globulus*. Honours thesis, University of Tasmania.

Chambers PGS, Potts BM, Tilyard PA (1997) The genetic control of flowering precocity in *Eucalyptus globulus* ssp. *globulus*. *Silvae Genetica* **46**, 207-214.

Cheema J, Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. *Briefings in Bioinformatics* **10**, 595-608.

Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L (2010) SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. *Journal of Integrative Plant Biology* **52**, 946-951.

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research* **15**, 1496-1502.

Climent J, Chambel MR, Lopez R, Mutke S, Alia R, Gil L (2006) Population divergence for heteroblasty in the Canary Island pine (*Pinus canariensis, Pinaceae*). *Am. J. Bot.* **93**, 840-848.

Collard B, Jahufer M, Brouwer J, Pang E (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169-196.

Collard B, Mace E, McPhail M, Wenzl P, Cakir M, Fox G, Poulsen D, Jordan D (2009) How accurate are the marker orders in crop linkage maps generated from large marker datasets? *Crop and Pasture Science* **60**, 362-372.

Cone KC, McMullen MD, *et al.* (2002) Genetic, physical, and informatics resources for *Maize*. On the road to an integrated map. *Plant Physiology* **130**, 1598-1605.

Cook IO, Ladiges PY (1991) Morphological variation within *Eucalyptus nitens* s. lat. and recognition of a new species, *E. denticulata*. *Australian Systematic Botany* **4**, 375-390.

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present day communities? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 1551-1571.

Daehler CC, Anttila CK, Ayres DR, Strong DR, Bailey JP (1999) Evolution of a new ecotype of *Spartina alterniflora* (*Poaceae*) in San Francisco Bay, California, USA. *American Journal of Botany* **86**, 543-546.

Doughty R (2000) 'The *Eucalyptus*: a natural and commercial history of the gum tree.' (John Hopkins University Press: Baltimore, MD)

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Dungey HS, Potts BM, Carnegie A, Ades PK (1997) *Mycosphaerella* leaf disease: genetic variation in damage to *Eucalyptus nitens*, *Eucalyptus globulus*, and their F_1 hybrid. *Canadian Journal of Forest Research* **27**, 750-759.

Dutkowski GW, Potts BM (1999) Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany* **47**, 237-263.

Dutkowski GW, Potts BM, Williams D, Kube P, McArthur C (2001) Geographic genetic variation in Central Victorian *Eucalyptus nitens*. In 'IUFRO Symposium on Developing the Eucalypt for the future'. Valdivia, Chile p. 39

Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. *American Journal of Botany* **93**, 953-962.

Eldridge KG, Davidson J, Harwood C, van Wyk G (1993) 'Eucalypt breeding and domestication.' (Clarendon Press: Oxford)

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Faria D, Mamani E, Pappas G, Grattapaglia D (2011) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. *Tree Genetics & Genomes* **7**, 63-77.

Farrow RA, Floyd RB, Neumann FG (1994) Inter-provenance variation in resistance of *Eucalyptus globulus* juvenile foliage to insect feeding. *Australian Forestry* **57**, 65-68.

Fernández-Ocaña A, Carmen García-López M, *et al.* (2010) Identification of a gene involved in the juvenile-to-adult transition (JAT) in cultivated olive trees. *Tree Genetics & Genomes* **6**, 891-903.

Ferreira A, Flores da Silva M, da Costa e Silva L, Cruz CD (2006) Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology* **29**, 187-192.

Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* **3**, 329-341.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE (2007) Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**, 370-380.

Freeman J, Whittock S, Potts B, Vaillancourt R (2009) QTL influencing growth and wood properties in *Eucalyptus globulus*. *Tree Genetics & Genomes* **5**, 713-722.

Freeman JS (2006) The application of molecular markers to *Eucalyptus globulus* tree improvement. PhD thesis, University of Tasmania.

Freeman JS, Potts BM, Shepherd M, Vaillancourt RE (2006) Parental and consensus linkage maps of *Eucalyptus globulus* using AFLP and microsatellite markers. *Silvae Genetica* **55**, 202-217.

Freeman JS, Potts BM, Vaillancourt RE (2008) Few Mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. *Genetics* **178**, 563-571.

Gan S, Shi J, Li M, Wu K, Wu J, Bai J (2003) Moderate-density molecular maps of *Eucalyptus urophylla* (S. T. Blake) and *E. tereticornis* (Smith) genomes based on RAPD markers. *Genetica* **118**, 59-67.

Gion J-M, Carouche A, *et al.* (2011) Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. *BMC Genomics* **12**:301.

Glaubitz JC, Emebiri LC, Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus sieberi*: Inheritance, diversity, and improved scoring of single-based differences. *Genome* **44**, 1041-1045.

Gore P, Potts BM (1995) The genetic control of flowering time in *Eucalyptus globulus*, *E. nitens* and their F_1 hybrids. In 'Eucalypt Plantations : Improving Fibre Yield and Quality - CRC-IUFRO Conference, 19 - 24 February 1995'. Hobart, Tasmania, Australia.

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* **144**, 1205-1214.

Grattapaglia D, Bradshaw HD (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research* **24**, 1074-1078.

Grattapaglia D, Kirst M (2008) Tansley review: *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist* **179**, 911-929.

Grattapaglia D, Plomion C, Kirst M, Sederoff RR (2009) Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* **12**, 148-156.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**, 1121-1137.

Grattapaglia D, Silva-Junior O, Kirst M, de Lima B, Faria D, Pappas G (2011) High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biology* **11:**65.

Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Kulheim C, Potts BM, Myburg AA (submitted) Progress in *Myrtaceae* genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics & Genomes*.

Gunn BV, McDonald MW (1991) *Eucalyptus urophylla* seed collections. *Forest Genetic Resources* **19**, 34-37 FAO, Rome.

Gupta P, Rustgi S, Mir R (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* **101**, 5-18.

Gustafson J, Ma X-F, Korzun V, Snape J (2009) A consensus map of rye integrating mapping data from five mapping populations. *Theoretical and Applied Genetics* **118**, 793-800.

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**, 33-38.

Hall MC, Willis JH (2006) Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution* **60**, 2466-2477.

Hamanishi ET, Campbell MM (2011) Genome-wide responses to drought in forest trees. *Forestry* **84**, 273-283.

Hamilton M, Potts B (2008) Review of *Eucalyptus nitens* genetic parameters. *New Zealand Journal of Forestry Science* **38**, 102-119.

Hamilton MG, Tilyard PA, Williams DR, Vaillancourt RE, Wardlaw TJ, Potts BM (2011) The genetic variation in the timing of heteroblastic transition in *Eucalyptus globulus* is stable across environments. *Australian Journal of Botany* **59**, 170-175.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research* **16**, 730-737.

Hasan O, Reid JB (1995) Reduction of generation time in *Eucalyptus globulus*. *Plant Growth Regulation* **17**, 53-60.

Hecht V, Foucher F, *et al.* (2005) Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiology* **137**, 1420-1434.

Henderson IR, Dean C (2004) Control of *Arabidopsis* flowering: the chill before the bloom. *Development* **131**, 3829-3838.

Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Sciences* **171**, 1059-1071.

Holderegger R, Herrmann D, *et al.* (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity* **1**, 273 - 283.

Hougaard BK, Madsen LH, *et al.* (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and Arachis. *Genetics* **179**, 2299-2312.

Hudson CJ, Kullan ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM, Vaillancourt RE (in press) High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. *Tree Genetics & Genomes*.

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss (*Buchloë dactyloides* (Nutt.) Engelm). *Theoretical and Applied Genetics* **86**, 927-934.

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution* **18**, 147-155.

Iglesias-Trabado G, Wilstermann D (2008) *Eucalyptus universalis*; global cultivated eucalypt forests map 2008 (Version 1.0.1). GIT Forestry Consulting - EUCALYPTOLOGICS. http://git-forestry.com/download_git_eucalyptus_map.htm. Accessed 1/5/2011.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**, (4):e25.

Jakobsson M, Rosenberg NA (2007) *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

James SA, Bell DT (2000) Leaf orientation, light interception and stomatal conductance of *Eucalyptus globulus* ssp. *globulus* leaves. *Tree Physiology* **20**, 815-823.

James SA, Bell DT (2001) Leaf morphological and anatomical characteristics of heteroblastic *Eucalyptus globulus* ssp. *globulus* (*Myrtaceae*). *Australian Journal of Botany* **49**, 259-269.

Jaya E, Kubien DS, Jameson PE, Clemens J (2010a) Vegetative phase change and photosynthesis in *Eucalyptus occidentalis*: architectural simplification prolongs juvenile traits. *Tree Physiology* **30**, 393-403.

Jaya E, Clemens J, Song J, Zhang H, Jameson PE (2010b) Quantitative expression analysis of meristem identity genes in *Eucalyptus occidentalis*: *AP1* is an expression marker for flowering. *Tree Physiology* **30**, 304-312.

Johansson AM, Pettersson ME, Siegel PB, Carlborg O (2010) Genome-wide effects of long-term divergent selection. *PLoS Genet* **6**, e1001188.

Jones M, Shepherd M, Henry R, Delves A (2006) Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics* **7**, 691-703.

Jones N, Ougham H, Thomas H (1997) Markers and mapping: we are all geneticists now. *New Phytologist* **137**, 165-177.

Jones N, Ougham H, Thomas H, Pašakinskienė I (2009) Markers and mapping revisited: finding your gene. *New Phytologist* **183**, 935-966.

Jones RC (2009) Molecular evolution and genetic control of flowering in the *Eucalyptus globulus* species complex. PhD thesis, University of Tasmania.

Jordan G, Borralho N, Tilyard P, Potts B (1994) Identification of races in *Eucalyptus globulus* ssp. *globulus* based on growth traits in Tasmania and geographic distribution. *Silvae Genetica* **43**, 292-298.

Jordan G, Potts BM, Wiltshire RJE (1999) Strong, independent, quantitative genetic control of the timing of vegetative phase change and first flowering in *Eucalyptus globulus* spp. *globulus* (Tasmanian Blue Gum). *Heredity* **83**, 179-187.

Jordan GJ, Potts BM, Chalmers P, Wiltshire RJE (2000) Quantitative genetic evidence that the timing of vegetative phase change in *Eucalyptus globulus* ssp. *globulus* is an adaptive trait. *Australian Journal of Botany* **48**, 561-567.

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology* **15**, 3469-3480.

Kaló P, Seres A, Taylor S, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis T, Kiss G (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular Genetics and Genomics* **272**, 235-246.

Kaya Z, Sewell MM, Neale DB (1999) Identification of quantitative trait loci influencing annual height and diameter-increment growth in loblolly pine (*Pinus taeda* L.). *Theoretical and Applied Genetics* **98**, 586-592.

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? *Heredity* **80**, 137-142.

Keats BJB, Sherman SL, *et al.* (1991) Guidelines for human linkage maps: An international system for human linkage maps (ISLM, 1990). *Genomics* **9**, 557-560.

Komeda Y (2004) Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**, 521-535.

Komulainen P, Brown GR, Mikkonen M, Karhu A, García-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theoretical and Applied Genetics* **107**, 667-678.

Kremer A, Caron H, *et al.* (2005) Monitoring genetic diversity in tropical trees with multilocus dominant markers. *Heredity* **95**, 274-280.

Kremer A, Casasoli M, *et al.* (2007) *Fagaceae* Trees. In 'Genome mapping and molecular breeding in plants'. (Ed. C Kole) pp. 161-188. (Springer-Verlag: Berlin)

Krishnan A, Guiderdoni E, *et al.* (2009) Mutant resources in rice for functional genomics of the grasses. *Plant Physiology* **149**, 165-170.

Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the *Pinaceae*. *Genetics* **168**, 447-461.

Kulheim C, Hui Yeoh S, Maintz J, Foley W, Moran G (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* **10:**452.

Kullan ARK, Mizrachi E, Jones N, Kanler K, Bayley A, Myburg AA (2010) A high-density consensus genetic linkage map of *Eucalyptus grandis*, *E. urophylla* and their F_1 hybrid based on DArT and microsatellite markers and an interspecific F_2 pseudo-backcross pedigree In 'Plant and Animal Genome 18 Conference'. San Diego, California. January 9-13. Abstract p212.

Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA (in press) High-density genetic linkage maps with over 2400 sequence-anchored DArT markers for genetic dissection in an F_2 pseudo-backcross of *Eucalyptus grandis* x *E. urophylla*. *Tree Genetics & Genomes*.

Kyozuka J, Harcourt R, Peacock WJ, Dennis ES (1997) *Eucalyptus* has functional equivalents of the *Arabidopsis AP1* gene. *Plant Molecular Biology* **35**, 573-584.

Ladiges PY, Udovicic F, Nelson G (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family *Myrtaceae*. *Journal of Biogeography* **30**, 989-998.

Lai J, Ma J, *et al.* (2004) Gene loss and movement in the Maize genome. *Genome Research* **14**, 1924-1931.

Laurie D, Devos K (2002) Trends in comparative genetics and their potential impacts on wheat and barley research. *Plant Molecular Biology* **48**, 729-740.

Lefebvre-Pautigny F, Wu F, *et al.* (2009) High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. *Tree Genetics & Genomes* **6**, 565-577.

Leinonen PH, Sandring S, Quilot B, Clauss MJ, Mitchell-Olds T, Agren J, Savolainen O (2009) Local adaptation in European populations of *Arabidopsis lyrata* (*Brassicaceae*). *Am. J. Bot.* **96**, 1129-1137.

Li H, Kilian A, Zhou M, Wenzl P, Huttner E, Mendham N, McIntyre L, Vaillancourt R (2010) Construction of a high-density composite map and comparative mapping of segregation distortion regions (SDRs) in barley. *Molecular Genetics and Genomics* **284**, 319-331.

Li H, Madden JL, Potts BM (1996) Variation in volatile leaf oils of the Tasmanian *Eucalyptus* species II. Subgenus *Symphyomyrtus*. *Biochemical Systematics and Ecology* **24**, 547-569.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* **27**, 237-277.

Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist* **183**, 776-788.

Lysak M, Koch M, Beaulieu J, Meister A, Leitch I (2009) The dynamic ups and downs of genome size evolution in *Brassicaceae*. *Mol Biol Evol* **26**, 85-98.

Mace E, Rami J-F, *et al.* (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biology* **9**, 1-14.

Marcar N, Crawford D (2004) 'Trees for saline landscapes.' Rural Industries Research and Development Corporation (RIRDC), Canberra, Australian Capital Territory.

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Molecular Ecology* **11**, 1145-1156.

Marques C, Brondani R, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics* **105**, 474-478.

Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, Liu BH, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theoretical and Applied Genetics* **96**, 727-737.

McDonald MW, Brooker MIH, Butcher PA (2009) A taxonomic revision of *Eucalyptus camaldulensis* (*Myrtaceae*). *Australian Systematic Botany* **22**, 257-285.

McKinney ML, McNamara KJ (1991) 'Heterochrony - The evolution of ontogeny.' (Plenum Press: New York)

Missiagia AA, Piacezzi AL, Grattapaglia D (2005) Genetic mapping of *Eef1*, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genetics & Genomes* **1**, 79-84.

Moran GF (1992) Patterns of genetic diversity in Australian tree species. *New Forests* **6**, 49-66.

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. *Current Opinion in Plant Biology* **10**, 149-155.

Munns R (2002) Comparative physiology of salt and water stress. *Plant, Cell & Environment* **25**, 239-250.

Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis, Eucalyptus globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach. *Theoretical and Applied Genetics* **107**, 1028-1042.

Myburg AA, Potts BM, Marques CM, Kirst M, Gion J, Grattapaglia D, Grima-Pettenatti J (2007) Eucalypts. In 'Genome mapping and molecular breeding in plants '. (Ed. C Kole) pp. 115-160. (Springer-Verlag: Berlin)

Neale D, Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* **11**, 149-155.

Neale DB, Krutovsky KV (2005) Comparative genetic mapping in trees: the group of conifers. In 'Biotechnology in agriculture and forestry, Vol. 55. Molecular marker systems in plant breeding and crop improvement'. (Eds H Lorz and G Wenzel) pp. 267-277. (Springer-Verlag Berlin)

Nielsen R, Bustamante C, *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**, e170.

Obeso JR (2002) The costs of reproduction in plants. *New Phytologist* **155**, 321-348.

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 185-205.

Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, Muleo R, Scarascia Mugnozza G, Sabatti M (2010) Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genetics & Genomes* **6**, 863-875.

Paterson AH, Bowers JE, *et al.* (2000) Comparative genomics of plant chromosomes. *Plant Cell* **12**, 1523-1540.

Paterson AH, Bowers JE, Feltus FA, Tang H, Lin L, Wang X (2009) Comparative genomics of grasses promises a bountiful harvest. *Plant Physiol.* **149**, 125-131.

Payn K, Dvorak W, Janse B, Myburg A (2008) Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*; endemic to seven islands in eastern Indonesia. *Tree Genetics & Genomes* **4**, 519-530.

Payn KG, Dvorak WS, Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (*Myrtaceae*). *Australian Journal of Botany* **55**, 673-683.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pelgas B, Beauseigle S, Acheré V, Jeandroz S, Bousquet J, Isabel N (2006) Comparative genome mapping among *Picea glauca*, *P. mariana* × *P. rubens* and *P. abies*, and correspondence with other *Pinaceae*. *Theoretical and Applied Genetics* **113**, 1371-1393.

Peng JH, Zadeh H, *et al.* (2004) Chromosome bin map of expressed sequence tags in homoeologous Group 1 of hexaploid wheat and homoeology with rice and *Arabidopsis*. *Genetics* **168**, 609-623.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* **31**, 62-64.

Perez-Figueroa A, Garcia A-Pereira MJ, Saura M, Roian-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. *Molecular Breeding* **7**, 275-291.

Poethig RS (2003) Phase change and the regulation of developmental timing in plants. *Science* **301**, 334-336.

Poke F, Vaillancourt RE, Potts B, Reid J (2005) Genomic research in *Eucalyptus*. *Genetica* **125**, 79-101.

Potts BM, McGowen MH, Williams DR, Suitor S, Jones TH, Gore PL, Vaillancourt RE (2008) Advances in reproductive biology and seed production systems of *Eucalyptus*: the case of *Eucalyptus globulus*. *Southern Forests: a Journal of Forest Science* **70**, 145-154.

Potts BM, Wiltshire RJE (1997) Eucalypt genetics and genecology. In 'Eucalypt ecology: individuals to ecosystems'. (Eds J Williams and J Woinarski) pp. 55-91. (Cambridge University Press: Cambridge)

Praça MM, Carvalho CR, Novaes CRDB (2009) Nuclear DNA content of three *Eucalyptus* species estimated by flow and image cytometry. *Australian Journal of Botany* **57**, 524-531.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* **20**, R208-R215.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* **436**, 793-800.

Reich PB, Wright IJ, Cavender-Bares J, Craine JM, Oleksyn J, Westoby M, Walters MB (2003) The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences* **164**, S143-S164.

Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci, and functional trait evolution in plants. *International Journal of Plant Sciences* **164**, S7-S20.

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. *Salmonidae*) species pairs. *Mol Biol Evol* **175**, 375-398.

Rosenberg NA (2004) *Distruct*: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Sambatti JBM, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* **60**, 696-710.

Sansaloni C, Petroli C, Carling J, Hudson C, Steane D, Myburg A, Grattapaglia D, Vaillancourt R, Kilian A (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* **6**, 16.

Sansaloni C, Petroli C, *et al.* (2009) High-density Diversity Arrays Technology (DArT) genotyping for cost-effective mapping and genome-wide selection in *Eucalyptus*. 'Plant & Animal Genome (PAG) XVII', January 10-14th, San Diego, USA. Abstract W24

Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**, 162-167.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**, 595-619.

Savolainen V, Anstett M-C, *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes C, Burg K, Kremer A (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615-1626.

Semagn K, Bjornstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology* **5**, 2569-2587.

Sewell M, Neale D (2000) Mapping quantitative traits in forest trees. In 'Molecular biology of woody plants Volume 1'. (Eds S Jain and S Minocha) pp. 407–423. (Kluwer Academic: Netherlands)

Shepherd M, Kasem S, Lee D, Henry R (2006) Construction of microsatellite linkage maps for *Corymbia*. *Silvae Genetica* **55**, 228-238.

Simpson GG, Gendall AR, Dean C (1999) When to switch to flowering. *Annual Review of Cell and Developmental Biology* **15**, 519-550.

Slate J (2008) Robustness of linkage maps in natural populations: a simulation study. *Proceedings of the Royal Society B* **275**, 695-702.

Southerton SG, Marshall H, Mouradov A, Teasdale RD (1998a) Eucalypt MADS-Box genes expressed in developing flowers. *Plant Physiology* **118**, 365-372.

Southerton SG, Strauss SH, Olive MR, Harcourt RL, Decroocq V, Zhu X, Llewellyn DJ, Peacock WJ, Dennis ES (1998b) *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. *Plant Molecular Biology* **37**, 897-910.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant Journal* **3**, 739-744.

Steane DA, Nicolle D, Potts BM (2007) Phylogenetic positioning of anomalous eucalypts by using ITS sequence data. *Australian Systematic Botany* **20**, 402-408.

Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, Vaillancourt RE (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (*Myrtaceae*) using high-throughput, genome-wide genotyping. *Molecular Phylogenetics and Evolution* **59**, 206-224.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Development and characterisation of microsatellite loci in *Eucalyptus globulus* (*Myrtaceae*). *Silvae Genetica* **50**, 89-91.

Steane DA, West AK, Potts BM, Ovenden JR, Reid JB (1991) Restriction fragment length polymorphisms in chloroplast DNA from six species of *Eucalyptus*. *Australian Journal of Botany* **39**, 399-414.

Steinbauer MJ (2002) Oviposition preference and neonate performance of *Mnesampela privata* in relation to heterophylly in *Eucalyptus dunnii* and *E. globulus*. *Agricultural and Forest Entomology* **4**, 245-253.

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Studer B, Kolliker R, *et al.* (2010) EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). *BMC Plant Biology* **10:**177.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* **320**, 486-488.

Teune J-H, Steger G (2010) NOVOMIR: De novo prediction of MicroRNA-Coding regions in a single plant-genome. *Journal of Nucleic Acids* **2010:** Article ID 495904

Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits. *Theoretical and Applied Genetics* **104**, 379-387.

Thumma B, Baltunis B, Bell J, Emebiri L, Moran G, Southerton S (2010) Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. *Tree Genetics & Genomes* **6**, 877-889.

Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in Cinnamoyl CoA Reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257-1265.

Tuskan GA, DiFazio S, *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596-1604.

Van Ooijen J (2006) JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V.: Wageningen, Netherlands.

Van Ooijen J, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Plant Research International: Wageningen, Netherlands.

Van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics* **84**, 803-811.

Van Ooijen JW (2009) MapQTL 6.0 - software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V.: Wageningen, Netherlands.

Varshney R, Marcel T, *et al.* (2007) A high density barley microsatellite consensus map with 775 SSR loci. *Theoretical and Applied Genetics* **114**, 1091-1103.

Velikova V, Loreto F, Brilli F, Stefanov D, Yordanov I (2008) Characterization of juvenile and adult leaves of *Eucalyptus globulus* showing distinct heteroblastic development: photosynthesis and volatile isoprenoids. *Plant Biology* **10**, 55-64.

Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, Kremer A (1997) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers: 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics* **95**, 597-608.

Verhoeven KJF, Poorter H, Nevo E, Biere A (2008) Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Molecular Ecology* **17**, 3416-3424.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *Heredity* **93**, 77-78.

Wang J-W, Park MY, Wang L-J, Koo Y, Chen X-Y, Weigel D, Poethig RS (2011) MiRNA control of vegetative phase change in trees. *PLoS Genet* **7**, e1002012.

Wei F, Coe E, *et al.* (2007) Physical and genetic structure of the *Maize* genome reflects its complex evolutionary history. *PLoS Genet* **3**, e123.

Wenzl P, Li H, *et al.* (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* **7:**206.

Wheeler NC, Jermstad KD, Krutovsky K, Aitken SN, Howe GT, Krakowski J, Neale DB (2005) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. *Molecular Breeding* **15**, 145-156.

Williams JE, Brooker MIH (1997) Eucalypts: an introduction. In 'Eucalypt ecology: individuals to ecosytems'. (Eds J Williams and J Woinarski) pp. 1-15. (Cambridge University Press: Cambridge U.K.)

Williams KJ, Potts BM (1996) The natural distribution of *Eucalyptus* species in Tasmania. *Tasforests* **8**, 39-165.

Wiltshire R, Murfet IC, Reid JB (1994) The genetic control of heterochrony: Evidence from developmental mutants of *Pisum sativum* L. *Journal of Evolutionary Biology* **7**, 447-465.

Wiltshire RJE, Reid JB, Potts BM (1998) Genetic control of reproductive and vegetative phase change in the *Eucalyptus risdonii* & *E. tenuiramis* complex. *Australian Journal of Botany* **46**, 45-63.

Wright IJ, Reich PB, Westoby M (2001) Strategy shifts in leaf physiology, structure and nutrient content between species of high- and low-rainfall and high- and low-nutrient habitats. *Functional Ecology* **15**, 423-434.

Wu F, Eannetta N, Xu Y, Durrett R, Mazourek M, Jahn M, Tanksley S (2009a) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theoretical and Applied Genetics* **118**, 1279-1293.

Wu G, Park MY, Conway SR, Wang J-W, Weigel D, Poethig RS (2009b) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* **138**, 750-759.

Wu G, Poethig RS (2006) Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* **133**, 3539-3547.

Yang L, Conway SR, Poethig RS (2011) Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development* **138**, 245-249.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8**, 907-913.

Zotz G, Wilhelm K, Becker A (2011) Heteroblasty - A Review. *The Botanical Review* **77**, 109-151.